(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/46256 A2

(51) International Patent Classification[7]: C07K 14/47

(21) International Application Number: PCT/US00/34919

(22) International Filing Date:
21 December 2000 (21.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/172,968	21 December 1999 (21.12.1999)	US
60/172,066	23 December 1999 (23.12.1999)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/172,968 (CIP)
Filed on	21 December 1999 (21.12.1999)
US	60/172,066 (CIP)
Filed on	23 December 1999 (23.12.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monrow Drive #17, Mountain View, CA 94040 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **BURFORD, Neil** [US/US]; 105 Wildwood Circle, Durham, CT 06522 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinely Avenue #3, Sunnyvale, CA 94086 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/46256 A2

(54) Title: VESICLE TRAFFICKING PROTEINS

(57) Abstract: The invention provides human vesicle trafficking proteins (VETRP) and polynucleotides which identify and encode VETRP. The invention also provides expession vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expession of VETRP.

VESICLE TRAFFICKING PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of vesicle trafficking proteins and to the use of these sequences in the diagnosis, treatment, and prevention of vesicle trafficking disorders, autoimmune/inflammatory disorders, and cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of vesicle trafficking proteins.

BACKGROUND OF THE INVENTION

Eukaryotic cells are bound by a lipid bilayer membrane and subdivided into functionally distinct, membrane-bound compartments. The membranes maintain the essential differences between the cytosol, the extracellular environment, and the lumenal space of each intracellular organelle. As lipid membranes are highly impermeable to most polar molecules, transport of essential nutrients, metabolic waste products, cell signaling molecules, macromolecules, and proteins across lipid membranes and between organelles must be mediated by a variety of transport-associated molecules.

Integral membrane proteins, secreted proteins, and proteins destined for the lumen of organelles are synthesized within the endoplasmic reticulum (ER), delivered to the Golgi complex for post-translational processing and sorting, and then transported to specific intracellular and extracellular destinations. Material is internalized from the extracellular environment by endocytosis, a process essential for transmission of neuronal, metabolic, and proliferative signals; uptake of many essential nutrients; and defense against invading organisms. This intracellular and extracellular movement of protein molecules is termed vesicle trafficking. Trafficking is accomplished by the packaging of protein molecules into specialized vesicles which bud from the donor organelle membrane and fuse to the target membrane (Rothman, J.E and F.T. Wieland (1996) Science 272:227-234).

The transport of proteins across the ER membrane involves a process that is similar in bacteria, yeast, and mammals (Gorlich, D. et al. (1992) Cell 71: 489-503). In mammalian systems, transport is initiated by the action of a cytoplasmic signal recognition particle (SRP) which recognizes a signal sequence on a growing, nascent polypeptide and binds the polypeptide and its ribosome complex to the ER membrane through an SRP receptor located on the ER membrane. The signal peptide is cleaved and the ribosome complex, together with the attached polypeptide, becomes membrane bound. The polypeptide is subsequently translocated across the ER membrane and into a vesicle (Blobel, G. and B. Dobberstein (1975) J. Cell Biol. 67:852-862).

Proteins implicated in the translocation of polypeptides across the ER membrane in yeast

include SEC61p, SEC62p, and SEC63p. Mutations in the genes encoding these proteins lead to defects in the translocation process. SEC61 may be of particular importance since certain mutations in the gene for this protein inhibit the translocation of many proteins (Gorlich, supra).

Mammalian homologs of yeast SEC61 (mSEC61) have been identified in dog and rat (Gorlich, supra). Mammalian SEC61 is also structurally similar to SECYp, the bacterial cytoplasmic membrane translocation protein. mSEC61 is found in tight association with membrane-bound ribosomes. This association is induced by membrane-targeting of nascent polypeptide chains and is weakened by dissociation of the ribosomes into their constituent subunits. mSEC61 is postulated to be a component of a putative protein-conducting channel, located in the ER membrane, to which nascent polypeptides are transferred following the completion of translation by ribosomes (Gorlich, supra).

Several steps in the transit of material along the secretory and endocytic pathways requires the formation of transport vesicles. Specifically, vesicles form at the transitional endoplasmic reticulum (tER), the rim of Golgi cisternae, the face of the Trans-Golgi Network (TGN), the plasma membrane (PM), and tubular extensions of the endosomes. Vesicle formation occurs when a region of membrane buds off from the donor organelle. The membrane-bound vesicle contains proteins to be transported and is surrounded by a proteinaceous coat, the components of which are recruited from the cytosol. Vesicle formation begins with the budding of a vesicle out of a donor organelle. The initial budding and coating processes are controlled by a cytosolic ras-like GTP-binding protein, ADP-ribosylating factor (Arf), and adapter proteins (AP). Different isoforms of both Arf and AP are involved at different sites of budding. For example, Arfs 1, 3, and 5 are required for Golgi budding, Arf4 for endosomal budding, and Arf6 for plasma membrane budding. Two different classes of coat protein have also been identified. Clathrin coats form on vesicles derived from the TGN and PM, whereas coatomer (COP) coats form on vesicles derived from the ER and Golgi (Mellman, I. (1996) Annu. Rev. Cell Dev. Biol. 12:575-625).

Vesicle formation begins when an adapter protein (AP) interacts with cargo proteins within the donor membrane and recruits clathrin to the bud site. APs are heterotetrameric complexes composed of two large chains (a, g, d, or e, and b), a medium chain (m), and a small chain (s). Clathrin binds to APs via the carboxy-terminal appendage domain of the b-adaptin subunit (Le Bourgne, R. and B. Hoflack (1998) Curr. Opin. Cell. Biol. 10:499-503). AP-1 functions in protein sorting from the TGN and endosomes to compartments of the endosomal/lysosomal system. AP-2 functions in clathrin-mediated endocytosis at the plasma membrane, while AP-3 is associated with endosomes and/or the TGN and recruits integral membrane proteins for transport to lysosomes and lysosome-related organelles. The recently isolated AP-4 complex localizes to the TGN or a neighboring compartment and may play a role in sorting events thought to take place in post-Golgi

compartments (Dell'Angelica, E.C. et al. (1999) J. Biol. Chem. 274:7278-7285). Cytosolic GTP-bound Arf is also incorporated into the vesicle as it forms. Another GTP-binding protein, dynamin, forms a ring complex around the neck of the forming vesicle and provides the mechanochemical force required to release the vesicle from the donor membrane. The coated vesicle complex is then transported through the cytosol. During the transport process, Arf-bound GTP is hydrolyzed to GDP and the coat dissociates from the transport vesicle (West, M.A. et al. (1997) J. Cell Biol. 138:1239-1254).

Coat protein (COP) coats form on the ER and Golgi. COP coats can further be distinguished as COPI, involved in retrograde traffic through the Golgi to the ER, and COPII, involved in anterograde traffic from the ER to the Golgi. The COP coat consists of two major components, a GTP-binding protein (Arf or Sar) and coat protomer (coatomer). Coatomer is an equimolar complex of seven proteins, termed alpha-, beta-, beta'-, gamma-, delta-, epsilon- and zeta-COP. The coatomer complex binds to dilysine motifs contained on the cytoplasmic tails of integral membrane proteins. These include the dilysine-containing retrieval motif of membrane proteins of the ER and dibasic/diphenylamine motifs of members of the p24 family. The p24 family of type I membrane proteins represent the major membrane proteins of COPI vesicles (Harter, C. and F.T. Wieland (1998) Proc. Natl. Acad. Sci. USA 95:11649-11654).

Vesicles can undergo homotypic or heterotypic fusion. Molecules required for appropriate targeting and fusion of vesicles include proteins in the vesicle membrane, the target membrane, and proteins recruited from the cytosol. During budding of the vesicle from the donor compartment, an integral membrane protein, VAMP (vesicle-associated membrane protein) is incorporated into the vesicle. Soon after the vesicle uncoats, a cytosolic prenylated GTP-binding protein, Rab, is inserted into the vesicle membrane. The amino acid sequence of Rab proteins reveals conserved GTP-binding domains characteristic of Ras superfamily members. In the vesicle membrane, GTP-bound Rab interacts with VAMP. Once the vesicle reaches the target membrane, a GTPase activating protein (GAP) in the target membrane converts the Rab protein to the GDP-bound form. A cytosolic protein, guanine-nucleotide dissociation inhibitor (GDI) then removes GDP-bound Rab from the vesicle membrane. Several Rab isoforms have been identified and appear to associate with specific compartments within the cell. For example, Rabs 4, 5, and 11 are associated with the early endosome, whereas Rabs 7 and 9 associate with the late endosome. These differences may provide selectivity in the association between vesicles and their target membranes (Novick, P. and M. Zerial (1997) Cur. Opin. Cell Biol. 9:496-504).

Docking of the transport vesicle with the target membrane involves the formation of a complex between the vesicle SNAP receptor (v-SNARE), target membrane (t-) SNAREs, and certain other membrane and cytosolic proteins. Many of these other proteins have been identified although

their exact functions in the docking complex remain uncertain (Tellam, J.T. et al. (1995) J. Biol. Chem. 270:5857-5863; Hata, Y. and T.C. Sudhof (1995) J. Biol. Chem. 270:13022-13028). N-ethylmaleimide sensitive factor (NSF) and soluble NSF-attachment protein (α-SNAP and β-SNAP) are two such proteins that are conserved from yeast to man and function in most intracellular membrane fusion reactions. Sec1 represents a family of yeast proteins that function at many different stages in the secretory pathway including membrane fusion. Recently, mammalian homologs of Sec1, called Munc-18 proteins, have been identified (Katagiri, H. et al. (1995) J. Biol. Chem. 270:4963-4966; Hata et al. supra).

The SNARE complex involves three SNARE molecules, one in the vesicular membrane and two in the target membrane. Together they form a rod-shaped complex of four α-helical coiled-coils. The membrane anchoring domains of all three SNAREs project from one end of the rod. This complex is similar to the rod-like structures formed by fusion proteins characteristic of the enveloped viruses, such as myxovirus, influenza, filovirus (Ebola), and the HIV and SIV retroviruses. (Skehel, J.J. and D.C. Wiley (1998) Cell 95:871-874). It has been proposed that the SNARE complex is sufficient for membrane fusion, suggesting that the proteins which associate with the complex provide regulation over the fusion event (Weber, T. et al. (1998) Cell 92:759-772). For example, in neurons, which exhibit regulated exocytosis, docked vesicles do not fuse with the presynaptic membrane until depolarization, which leads to an influx of calcium (Bennett, M.K. and R.H. Scheller (1994) Annu. Rev. Biochem. 63:63-100). Synaptotagmin, an integral membrane protein in the synaptic vesicle, associates with the t-SNARE syntaxin in the docking complex. Synaptotagmin binds calcium in a complex with negatively charged phospholipids, which allows the cytosolic SNAP protein to displace synaptotagmin from syntaxin and fusion to occur. Thus, synaptotagmin is a negative regulator of fusion in the neuron (Littleton, J.T. et al. (1993) Cell 74:1125-1134). The most abundant membrane protein of synaptic vesicles appears to be the glycoprotein synaptophysin,, a 38 kDa protein with four transmembrane domains. Although the function of synaptophysin is not known, its calcium-binding ability, tyrosine phosphorylation, and widespread distribution in neural tissues suggest a potential role in neurosecretion (Bennett, supra).

The transport of proteins into and out of vesicles relies on interactions between cell membranes and a supporting membrane cytoskeleton consisting of spectrin and other proteins. A large family of related proteins called ankyrins participate in the transport process by binding to the membrane skeleton protein spectrin and to a protein in the cell membrane called band 3, a component of an anion channel in the cell membrane. Ankyrins therefore function as a critical link between the cytoskeleton and the cell membrane.

Originally found in association with erythroid cells, ankyrins are also found in other tissues as well (Birkenmeier, C.S. et al. (1993) J. Biol. Chem. 268:9533-9540). Ankyrins are large proteins

(~1800 amino acids) containing an N-terminal, 89 kDa domain that binds the cell membrane proteins band 3 and tubulin, a central 62 kDa domain that binds the cytoskeletal proteins spectrin and vimentin, and a C-terminal, 55 kDa regulatory domain that functions as a modifier of the binding activities of the other two domains. Individual genes for ankyrin are able to produce multiple ankyrin isoforms by various insertions and deletions. These isoforms are of nearly identical size but may have different functions. In addition, smaller transcripts are produced which are missing large regions of the coding sequences from the N-terminal (band 3 binding), and central (spectrin binding) domains. The existence of such a large family of ankyrin proteins and the observation that more than one type of ankyrin may be expressed in the same cell type suggests that ankyrins may have more specialized functions than simply binding the membrane skeleton to the plasma membrane (Birkenmeier, supra).

In humans, two isoforms of ankyrin are expressed, alternatively, in developing erythroids and mature erythroids, respectively (Lambert, S. et. al. (1990) Proc. Natl. Acad. Sci. USA 87:1730-1734). A deficiency in erythroid spectrin and ankyrin has been associated with the hemolytic anemia, hereditary spherocytosis (Coetzer, T.L. et al. (1988) New Engl. J. Med. 318:230-234).

Correct trafficking of proteins is of particular importance for the proper function of epithelial cells, which are polarized into distinct apical and basolateral domains containing different cell membrane components such as lipids and membrane-associated proteins. Certain proteins are flexible and may be sorted to the basolateral or apical side depending upon cell type or growth conditions. For example, the kidney anion exchanger (kAE1) can be retargeted from the apical to the basolateral domain if cells are plated at higher density. The protein kanadaptin was isolated as a protein which binds to the cytoplasmic domain of kAE1. It also colocalizes with kAE1 in vesicles, but not in the membrane, suggesting that kanadaptin's function is to guide kAE1-containing vesicles to the basolateral target membrane (Chen, J. et al. (1998) J. Biol. Chem. 273:1038-1043).

Vesicle trafficking is crucial in the process of neurotransmission. Synaptic vesicles carry neurotransmitter molecules from the cytoplasm of a neuron to the synapse. Rab3's are a family of GTP-binding proteins located on synaptic vesicles. The RIM family of proteins are thought to be effectors for Rab3's (Wang, Y. et al. (2000) J. Biol. Chem. 275:20033-20044). Rabphilin-3 is a synaptic vesicle protein. Granuphilins are proteins with homology to rabphilins, and may have a unique role in exocytosis (Wang, J. et al. (1999) J. Biol. Chem. 274:28542-28548).

The etiology of numerous human diseases and disorders can be attributed to defects in the trafficking of proteins to organelles or the cell surface. Defects in the trafficking of membrane-bound receptors and ion channels are associated with cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR), glucose-galactose malabsorption syndrome (Na^+/glucose cotransporter), hypercholesterolemia (low-density lipoprotein (LDL) receptor), and forms of diabetes

mellitus (insulin receptor). Abnormal hormonal secretion is linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), and Cushing's and Addison's diseases (adrenocorticotropic hormone; ACTH).

Cancer cells secrete excessive amounts of hormones or other biologically active peptides. Disorders related to excessive secretion of biologically active peptides by tumor cells include: fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors; hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia; and carcinoid syndrome, which includes abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Ectopic synthesis and secretion of biologically active peptides (peptides not expected from a tumor) includes ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating hormone in medullary thyroid carcinoma.

Various human pathogens alter host cell protein trafficking pathways to their own advantage. For example, the HIV protein Nef downregulates cell-surface expression of CD4 molecules by accelerating their endocytosis through clathrin coated pits. This function of Nef is important for the spread of HIV from the infected cell (Harris, M. (1999) Curr. Biol. 9:R449-R461). A recently identified human protein, Nef-associated factor 1 (Naf1), a protein with four extended coiled-coil domains, has been found to associate with Nef. Overexpression of Naf1 increased cell surface expression of CD4, an effect which could be suppressed by Nef (Fukushi, M. et al. (1999) FEBS Lett. 442:83-88).

The discovery of new vesicle trafficking proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of vesicle trafficking disorders, autoimmune/inflammatory disorders, and cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of vesicle trafficking proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, vesicle trafficking proteins, referred to collectively as "VETRP" and individually as "VETRP-1," "VETRP-2," "VETRP-3," "VETRP-4," "VETRP-5," "VETRP-6," "VETRP-7," "VETRP-8," "VETRP-9," "VETRP-10," "VETRP-11," "VETRP-12," "VETRP-13," "VETRP-14," "VETRP-15," "VETRP-16," "VETRP-17," "VETRP-18," "VETRP-19," "VETRP-20," "VETRP-21," "VETRP-22," and "VETRP-23." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the

group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-23.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-23. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:24-46.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the

polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said

target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional VETRP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional VETRP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The

method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional VETRP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:24-46, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding VETRP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of VETRP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding VETRP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and

polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"VETRP" refers to the amino acid sequences of substantially purified VETRP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of VETRP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of VETRP either by directly interacting with VETRP or by acting on components of the biological pathway in which VETRP participates.

An "allelic variant" is an alternative form of the gene encoding VETRP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding VETRP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as VETRP or a polypeptide with at least one functional characteristic of VETRP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding VETRP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding VETRP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent VETRP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of VETRP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of VETRP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of VETRP either by directly interacting with VETRP or by acting on components of the biological pathway in which VETRP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind VETRP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or

oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic VETRP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding VETRP or fragments of VETRP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be

associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the

absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of VETRP or the polynucleotide encoding VETRP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:24-46 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:24-46, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:24-46 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:24-46 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:24-46 and the region of SEQ ID NO:24-46 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-23 is encoded by a fragment of SEQ ID NO:24-46. A fragment of SEQ ID NO:1-23 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-23. For example, a fragment of SEQ ID NO:1-23 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-23. The precise length of a fragment of SEQ ID NO:1-23 and the region of SEQ ID NO:1-23 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended

purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and

conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of VETRP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of VETRP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of VETRP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of VETRP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an VETRP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of VETRP.

"Probe" refers to nucleic acid sequences encoding VETRP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have

been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding VETRP, or fragments thereof, or VETRP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides

by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example,

an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human vesicle trafficking proteins (VETRP), the polynucleotides encoding VETRP, and the use of these compositions for the diagnosis, treatment, or prevention of vesicle trafficking disorders, autoimmune/inflammatory disorders, and cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding VETRP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each VETRP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each VETRP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6

shows homologous sequences as identified by BLAST analysis along with relevant citations, all of which are expressly incorporated by reference herein in their entirety; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding VETRP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:24-46 and to distinguish between SEQ ID NO:24-46 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express VETRP as a fraction of total tissues expressing VETRP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing VETRP as a fraction of total tissues expressing VETRP. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:25 in nervous tissue. SEQ ID NO:41 is noted for its expression in both cancer and reproductive tissue, and SEQ ID NO:43 is expressed in cancer and nervous tissue.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding VETRP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses VETRP variants. A preferred VETRP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the VETRP amino acid sequence, and which contains at least one functional or structural characteristic of VETRP.

SEQ ID NO:31 maps to chromosome 12 within the interval from 70.60 to 76.50 centiMorgans, and to chromosome 1 within the interval from 159.60 to 164.10 centiMorgans. SEQ ID NO:36 maps to chromosome 3 within the interval from 129.00 to 131.80 centiMorgans, and to chromosome 4 within the interval from 86.00 to 91.90 centiMorgans. SEQ ID NO:38 maps to chromosome 6 within the interval from the p-terminus to 27.10 centiMorgans. SEQ ID NO:42 maps to chromosome 2 within the interval from 233.10 to 236.10 centiMorgans. SEQ ID NO:44 maps to chromosome 5 within the interval from 61.10 to 69.60 centiMorgans, to chromosome 11 within the interval from 117.90 to 123.50 centiMorgans, and to chromosome 17 within the interval from 99.30 to 103.70 centiMorgans.

The invention also encompasses polynucleotides which encode VETRP. In a particular

embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46, which encodes VETRP. The polynucleotide sequences of SEQ ID NO:24-46, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding VETRP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding VETRP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:24-46. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of VETRP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding VETRP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring VETRP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode VETRP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring VETRP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding VETRP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding VETRP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode VETRP and VETRP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell

systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding VETRP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:24-46 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding VETRP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region

of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode VETRP may be cloned in recombinant DNA molecules that direct expression of VETRP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express VETRP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter VETRP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of VETRP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding VETRP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, VETRP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of VETRP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active VETRP, the nucleotide sequences encoding VETRP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers,

constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding VETRP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding VETRP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding VETRP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding VETRP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding VETRP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for

delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding VETRP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding VETRP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding VETRP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of VETRP are needed, e.g. for the production of antibodies, vectors which direct high level expression of VETRP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of VETRP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of VETRP. Transcription of sequences encoding VETRP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding VETRP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain

infective virus which expresses VETRP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of VETRP in cell lines is preferred. For example, sequences encoding VETRP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the

sequence encoding VETRP is inserted within a marker gene sequence, transformed cells containing sequences encoding VETRP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding VETRP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding VETRP and that express VETRP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of VETRP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on VETRP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding VETRP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding VETRP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding VETRP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence

and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode VETRP may be designed to contain signal sequences which direct secretion of VETRP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding VETRP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric VETRP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of VETRP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the VETRP encoding sequence and the heterologous protein sequence, so that VETRP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled VETRP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

VETRP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to VETRP. At least one and up to a plurality of test compounds may be screened for specific binding to VETRP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of VETRP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which VETRP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express VETRP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing VETRP or cell membrane fractions which contain VETRP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either VETRP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with VETRP, either in solution or affixed to a solid support, and detecting the binding of VETRP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

VETRP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of VETRP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for VETRP activity, wherein VETRP is combined with at least one test compound, and the activity of VETRP in the presence of a test compound is compared with the activity of VETRP in the absence of the test compound. A change in the activity of VETRP in the presence of the test compound is indicative of a compound that modulates the activity of VETRP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising VETRP under conditions suitable for VETRP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of VETRP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding VETRP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding VETRP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding VETRP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding VETRP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress VETRP, e.g., by secreting VETRP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of VETRP and vesicle trafficking proteins. In addition, the expression of VETRP is closely associated with reproductive tissue, nervous tissue, cancer and inflammation/trauma.

Therefore, VETRP appears to play a role in vesicle trafficking disorders, autoimmune/inflammatory disorders, and cancer. In the treatment of disorders associated with increased VETRP expression or activity, it is desirable to decrease the expression or activity of VETRP. In the treatment of disorders associated with decreased VETRP expression or activity, it is desirable to increase the expression or activity of VETRP.

Therefore, in one embodiment, VETRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of VETRP. Examples of such disorders include, but are not limited to, a vesicle trafficking disorder, such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminthic, and protozoal infections; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cancer, such as, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and

uterus.

In another embodiment, a vector capable of expressing VETRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of VETRP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified VETRP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of VETRP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of VETRP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of VETRP including, but not limited to, those listed above.

In a further embodiment, an antagonist of VETRP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of VETRP. Examples of such disorders include, but are not limited to, those vesicle trafficking disorders, autoimmune/inflammatory disorders, and cancer described above. In one aspect, an antibody which specifically binds VETRP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express VETRP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding VETRP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of VETRP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of VETRP may be produced using methods which are generally known in the art. In particular, purified VETRP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind VETRP. Antibodies to VETRP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans,

and others may be immunized by injection with VETRP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to VETRP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of VETRP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to VETRP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce VETRP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for VETRP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and

easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between VETRP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering VETRP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for VETRP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of VETRP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple VETRP epitopes, represents the average affinity, or avidity, of the antibodies for VETRP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular VETRP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the VETRP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of VETRP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of VETRP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding VETRP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene

encoding VETRP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding VETRP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding VETRP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in VETRP expression or regulation causes disease, the expression of VETRP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in VETRP are treated by constructing mammalian expression vectors encoding VETRP and introducing these vectors by mechanical means into VETRP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of VETRP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). VETRP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding VETRP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to VETRP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding VETRP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc.

Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding VETRP to cells which have one or more genetic abnormalities with respect to the expression of VETRP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding VETRP to target cells which have one or more genetic abnormalities with respect to the expression of VETRP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing VETRP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a

cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding VETRP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for VETRP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of VETRP-coding RNAs and the synthesis of high levels of VETRP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of VETRP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of

mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding VETRP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding VETRP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding VETRP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide

sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased VETRP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding VETRP may be therapeutically useful, and in the treament of disorders associated with decreased VETRP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding VETRP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding VETRP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding VETRP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding VETRP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of VETRP, antibodies to VETRP, and mimetics, agonists, antagonists, or inhibitors of VETRP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising VETRP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, VETRP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example VETRP or fragments thereof, antibodies of VETRP, and agonists, antagonists or inhibitors of VETRP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind VETRP may be used for the diagnosis of disorders characterized by expression of VETRP, or in assays to monitor patients being treated with VETRP or agonists, antagonists, or inhibitors of VETRP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics.

Diagnostic assays for VETRP include methods which utilize the antibody and a label to detect VETRP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring VETRP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of VETRP expression. Normal or standard values for VETRP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to VETRP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of VETRP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding VETRP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of VETRP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of VETRP, and to monitor regulation of VETRP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding VETRP or closely related molecules may be used to identify nucleic acid sequences which encode VETRP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding VETRP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the VETRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:24-46 or from genomic sequences including promoters, enhancers, and introns of the VETRP gene.

Means for producing specific hybridization probes for DNAs encoding VETRP include the cloning of polynucleotide sequences encoding VETRP or VETRP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA

polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding VETRP may be used for the diagnosis of disorders associated with expression of VETRP. Examples of such disorders include, but are not limited to, a vesicle trafficking disorder, such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminthic, and protozoal infections; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cancer, such as, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding VETRP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect

altered VETRP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding VETRP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding VETRP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding VETRP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of VETRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding VETRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding VETRP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding VETRP, or a fragment of a polynucleotide complementary to the polynucleotide encoding

VETRP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding VETRP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding VETRP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of VETRP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to

monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for VETRP, or VETRP or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes

are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be

obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for VETRP to quantify the levels of VETRP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al.

(1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding VETRP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding VETRP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, VETRP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between VETRP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with VETRP, or fragments thereof, and washed. Bound VETRP is then detected by methods well known in the art. Purified VETRP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding VETRP specifically compete with a test compound for binding VETRP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with VETRP.

In additional embodiments, the nucleotide sequences which encode VETRP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/172,968 and U.S. Ser. No. 60/172,066 are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged

over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence

scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire

annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:24-46. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the

entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding VETRP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of ABBR Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:24-46 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:24-46 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:31, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:42, and SEQ ID NO:44 are described in The Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:31, SEQ ID NO:36, and SEQ ID NO:44, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:31, SEQ ID NO:36, and SEQ ID NO:44 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web

site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VI. Extension of VETRP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:24-46 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI-B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%)

agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:24-46 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:24-46 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37 °C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85 °C to the stop the reaction and degrade the RNA. Samples are purified

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is

incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each

spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the VETRP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring VETRP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of VETRP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the VETRP-encoding transcript.

X. Expression of VETRP

Expression and purification of VETRP is achieved using bacterial or virus-based expression systems. For expression of VETRP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express VETRP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of VETRP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding VETRP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, VETRP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on

immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from VETRP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified VETRP obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of VETRP Activity

VETRP activity is measured by its inclusion in coated vesicles. VETRP can be expressed by transforming a mammalian cell line such as COS7, HeLa, or CHO with an eukaryotic expression vector encoding VETRP. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes, such as β-galactosidase, is co-transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of VETRP and β-galactosidase.

Transformed cells are collected and cell lysates are assayed for vesicle formation. A non-hydrolyzable form of GTP, GTPγS, and an ATP regenerating system are added to the lysate and the mixture is incubated at 37 °C for 10 minutes. Under these conditions, over 90% of the vesicles remain coated (Orci, L. et al (1989) Cell 56:357-368). Transport vesicles are salt-released from the Golgi membranes, loaded under a sucrose gradient, centrifuged, and fractions are collected and analyzed by SDS-PAGE. Co-localization of VETRP with clathrin or COP coatamer is indicative of VETRP activity in vesicle formation. The contribution of VETRP in vesicle formation can be confirmed by incubating lysates with antibodies specific for VETRP prior to GTPγS addition. The antibody will bind to VETRP and interfere with its activity, thus preventing vesicle formation.

In the alternative, VETRP activity is measured by its ability to alter vesicle trafficking pathways. Vesicle trafficking in cells transformed with VETRP is examined using fluorescence microscopy. Antibodies specific for vesicle coat proteins or typical vesicle trafficking substrates such as transferrin or the mannose-6-phosphate receptor are commercially available. Various cellular components such as ER, Golgi bodies, peroxisomes, endosomes, lysosomes, and the plasmalemma are examined. Alterations in the numbers and locations of vesicles in cells transformed with VETRP as compared to control cells are characteristic of VETRP activity.

XII. Functional Assays

VETRP function is assessed by expressing the sequences encoding VETRP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of VETRP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding VETRP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding VETRP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of VETRP Specific Antibodies

VETRP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the VETRP amino acid sequence is analyzed using LASERGENE software

(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-VETRP activity by, for example, binding the peptide or VETRP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring VETRP Using Specific Antibodies

Naturally occurring or recombinant VETRP is substantially purified by immunoaffinity chromatography using antibodies specific for VETRP. An immunoaffinity column is constructed by covalently coupling anti-VETRP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing VETRP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of VETRP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/VETRP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and VETRP is collected.

XV. Identification of Molecules Which Interact with VETRP

VETRP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled VETRP, washed, and any wells with labeled VETRP complex are assayed. Data obtained using different concentrations of VETRP are used to calculate values for the number, affinity, and association of VETRP with the candidate molecules.

Alternatively, molecules interacting with VETRP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

VETRP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions

between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	24	381039	HYPONOB01	381039H1 (HYPONOB01), 1558030F6 (BLADTUT04), 1653756H1 (PROSTUT08), 2327916X32C1 (COLNNOT11), 2327916X57C1 (COLNNOT11), 2503308H1 (CONUTUT01), 2649725F6 (KIDNFET01), 3105456F6 (HEAONOT05), 3278946H1 (STOMFET02)
2	25	383249	HYPONOB01	383249H1 (HYPONOB01), 4310090T6 (BRAUNOT01), SXBC01838V1, SXBC00070V1, SCSA04368V1, SXBC00956V1
3	26	618769	PGANNOT01	618769H1 (PGANNOT01), 618769R6 (PGANNOT01), 897423R1 (BRSTNOT05), 897423T1 (BRSTNOT05), 2225425T6 (SEMVNOT01), 3026516F6 (HEARFET02), 3031626H2 (TLYMNOT05)
4	27	1234837	LUNGFET03	763436R1 (BRAITUT02), 1234837H1 (LUNGFET03), 1237053F6 (LUNGFET03), 1721329F6 (BLADNOT06), 2751788R7 (THP1AZS08), g777187
5	28	1607223	LUNGNOT15	1250975F6 (LUNGFET03), 1270831F1 (TESTTUT02), 1399306F1 (BRAITUT08), 1607223H1 (LUNGNOT15), 1684804F6 (PROSNOT15), 1803848F6 (SINTNOT13), 2814466T6 (OVARNOT10), 3274147F6 (PROSBPT06), 3289542F6 (BONRFET01), SCFA04890V1
6	29	1621554	BRAITUT13	795816F1 (OVARNOT03), 795816R1 (OVARNOT03), 855789R1 (NGANNOT01), 1357564T6 (LUNGNOT09), 1621554H1 (BRAITUT13), 2081463H1 (UTRSNOT08), 2474413F6 (SMCANOT01), 2812794F6 (OVARNOT10), 3296847H1 (TLYJINT01), 4642794H1 (PROSTMT03)
7	30	1751553	LIVRTUT01	816693R6 (OVARTUT01), 816693X311D1 (OVARTUT01), 816693X313D1 (OVARTUT01), 1678292F6 (STOMFET01), 1678292T6 (STOMFET01), 1751553H1 (LIVRTUT01), 1981902R6 (LUNGTUT03), 1981902T6 (LUNGTUT03), 3050018H1 (LUNGNOT25), 4419520T6 (LIVRDIT02), 5208961H1 (BRAFNOT02), g2823700
8	31	1832403	BRAINON01	025882F1 (SPLNFET01), 1255756F2 (MENITUT03), 1832403H1 (BRAINON01), 2305321R6 (NGANNOT01)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
9	32	1971747	UCMCL5T01	285815F1 (EOSIHET02), 936208R1 (CERVNOT01), 1510143H1 (LUNGNOT14), 1673683F6 (BLADNOT05), 1912830F6 (LEUKNOT02), 1971747F6 (UCMCL5T01), 1971747H1 (UCMCL5T01), 2264823R6 (UTRSNOT02), 2858746F6 (SININOT03), 3108785F6 (BRSTTUT15), 3391023H1 (LUNGTUT17), 3736985F6 (SMCCNOS01), SBVA02967V1, SBVA04682V1, SBVA04527V1, g3882214
10	33	2285348	BRAINON01	228256F1 (PANCNOT01), 453676H1 (TLYMNOT02), 857831R1 (NGANNOT01), 1350071F1 (LATRTUT02), 1558538F1 (SPLNNOT04), 2170558F6 (ENDCNOT03), 2285348H1 (BRAINON01), 2291739T6 (BRAINON01), 2418450F6 (HNT3AZT01), 2717843F6 (THYRNOT09), 3404595F6 (ESOGNOT03), 3427632F6 (BRSTNOR01)
11	34	2374186	ISLTNOT01	1309077F1 (COLNFET02), 1526217F1 (UCMCL5T01), 2374186H1 (ISLTNOT01), 3581348F6 (293TF3T01), 3581348T6 (293TF3T01), SCGA06229V1, SCGA12945V1, SCGA02017V1, SCGA13028V1
12	35	2476232	SMCANOT01	1394147T1 (THYRNOT03), 1709170F6 (PROSNOT16), 2476232F6 (SMCANOT01), 2476232H1 (SMCANOT01), 2733026T6 (OVARTUT04), 3589738H1 (293TF5T01), 4638417T6 (MYEPTXT01), 5027954H1 (COLCDIT01)
13	36	2503986	CONUTUT01	449043X14 (TLYMNOT02), 632170R6 (KIDNNOT05), 1450259F1 (PENITUT01), 1798335F6 (COLNNOT27), 2503986H1 (CONUTUT01), 2872984H1 (THYRNOT10), 3004241T6 (TLYMNOT06), 3027284T6 (HEARFET02), SBFA03755F1, SBFA04352F1, SBFA04697F1, SBFA00555F1, SBFA01887F1
14	37	2596566	OVARTUT02	349546R6 (LVENNOT01), 840332R1 (PROSTUT05), 1333240F6 (COLNNOT13), 1628086F6 (COLNPOT01), 3293723F6 (TLYJINT01), 4008413H1 (ENDCNOT04), 4818893H1 (PROSTUT17), 2596566H2 (OVARTUT02)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
15	38	2685253	LUNGNOT23	680657R6 (UTRSNOT02), 736092R1 (TONSNOT01), 1486268T6 (CORPNOT02), 2445454T6 (THP1NOT03), 2685253H1 (LUNGNOT23), 3731545H1 (SMCCNON03), 4020957H1 (BRAXNOT02), 4741584H1 (THYMNOR02), SZAT00201V1, SZAT01557V1, SZAT01529V1, SZAT00077V1, SZAT01181V1, SZAT01587V1, SZAT00683V1, SZAT00014V1, SZAT00017V1
16	39	2762252	BRSTNOT12	2685254H1 (LUNGNOT23), 2762252H1 (BRSTNOT12), 2766358F6 (BRSTNOT12), 2766358T6 (BRSTNOT12), 4364345H1 (SKIRNOT01), g3560562
17	40	3452009	UTRSNON03	3149588T6 (ADRENON04), 3452009H1 (UTRSNON03), SBHA03400F1, SBHA02544F1, SBHA03511F1
18	41	4644780	PROSTMT03	1494403R6 (PROSNON01), 1818377F6 (PROSNOT20), 3362284H1 (PROSBPT02), 3364760T6 (PROSBPT02), 4643280H1 (PROSTMT03), 4644736H1 (PROSTMT03), 4815885H1 (PROSTUS11), 5423826H1 (PROSTMT07), 5424290H1 (PROSTMT07)
19	42	4946103	SINTNOT25	1482075H1 (CORPNOT02), 1729337F6 (BRSTTUT08), 1729337X14C1 (BRSTTUT08), 1729337X16C1 (BRSTTUT08), 1730230X11C1 (BRSTTUT08), 1899544F6 (BLADTUT06), 2290603R6 (BRAINON01), 3042988H1 (HEAANOT01), 4946103H1 (SINTNOT25), SBAA00190F1, SBAA03996F1
20	43	5562355	BRSTDIT01	1909929F6 (CONNTUT01), 4027671H1 (BRAINOT23), 5091054F6 (UTRSTMR01), 5512655H1 (BRADDIR01), 5562355H1 (BRSTDIT01), g1625519
21	44	5678824	BRAENOT02	320465F1 (EOSIHET02), 506242X23R1 (TMLR3DT02), 1255854F2 (MENITUT03), 1351588F6 (LATRTUT02), 2051796F6 (LIVRFET02), 3421889H1 (UCMCNOT04), 5678824H1 (BRAENOT02), SCAA06386V1, SCAA02158V1, SBYA05287U1, SCAA04949V1

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
22	45	5870962	COLTDIT04	1415307F6 (BRAINOT12), 1504159F6 (BRAITUT07), 1692553T6 (COLNNOT23), 1868556T6 (SKINBIT01), 2911311H1 (KIDNTUT15), 2912463F6 (KIDNTUT15), 2926162T6 (TLYMNOT04), 5121545H1 (SMCBUNT01), 5812063H1 (KIDCTMT02), 5870962H1 (COLTDIT04), 6094909H1 (THP1TXT03)
23	46	2818605		2818605F6 (BRSTNOT14), 4904212T6 (TLYMNOT08), 1495645T6 (PROSNON01)

Table 2

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
1	481	S222 T56 T90 T92 S168 S313 S436 S217 S442	N433	Putative GTPase activating protein for ADP ribosylation factor (ArfGAP): C30-G153 HIV REV interacting protein: N44-P80, V83-N104, N43-W144 GATA-type zinc finger domain: N44-E93	g3135319 nucleoporin (Homo sapiens) Mikoshiba,K. et al. (1999) Chem. Phys. Lipids 98:59-67; Salcini,A.E. et al. (1997) Genes Dev. 11:2239-2249; Glockner,G. et al. (1998) Genome Res. 8: 1060-1073	Motifs BLAST-GENBANK BLIMPS_PRINTS BLAST_DOMO HMMER_PFAM BLIMPS_PRODOM
2	195	S40 S110 S144		Podocalyxin-like protein: M1-T195 P24 protein: M1-T195	g1890141 Neural organelle transport protein P24 (Mus musculus) Kadota,Y. et al. (1997) Brain Res. Mol. Brain Res. 46:265-273	Motifs BLAST-GENBANK BLAST_PRODOM

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
3	313	T66 T76 S150 T212 S251 T256 S277 T80 S130 T155	N8	Protein kinase C C2 domain: L26-I115, L185-T273 Protein kinase C C2 domain: I13-Q69 C2 domain motif: L54-E79 C2 domain: G11-T128	g5926736 granuphilin-a (Mus musculus) Wang,J. et al. (1999) J. Biol. Chem. 274: 28542-28548	Motifs BLAST-GENBANK HMMER_PFAM PROFILESCAN BLIMPS_PFAM BLAST_DOMO
4	201	S11 T14 S82 T94 S119 T149		SNF7 Nuclear transcription regulation protein motif: N3-E166	g3873551 coiled-coil protein (Schizosaccharomyces pombe)	Motifs BLAST-GENBANK BLAST_PRODOM
5	566	S112 S171 S195 S250 S282 S345 T381 T443 S503 T542 T547 S26 S35 S139 T377 S433 Y47	N213 N387	Gamma-adaptin Clathrin assembly protein complex: S407-W563 Gamma adaptin motif: M331-F541	g961444 related to mouse gamma adaptin (Homo sapiens) Nagase,T. et al. (1995) DNA Res. 2: 167-174	Motifs BLAST-GENBANK BLAST_PRODOM BLAST_DOMO
6	270	S192 S34 T91 S131 T146 T266 T18 T50 S187 S246			g4689260 sorting nexin 10 (Homo sapiens) Kurten,R. et al. (1996) Science 272:1008-1010	Motifs BLAST-GENBANK

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
7	490	S205 T320 T407 S94 S106 S156 S174 S218 S235 S255 S279 T320 S332 S397 S430 T18 S50 S102 S141 T162 S182 S200 S201 S208 S263 T304 T315 S327 S445	N22 N38 N180 N209 N234	C2 domain: I368-W456	g2724126 synaptotagmin VII (Homo sapiens) Cooper,P.R. et al. (1998) Genomics 49: 419-429	Motifs BLAST-GENBANK HMMER_PFAM
8	136	T20 S50 T20 T117 Y81 Y129	N110		g4206090 snapin SNARE-associated synaptic transmission protein (Mus musculus) Ilardi,J.M. et al. (1999) Nat. Neurosci. 2:119-124	Motifs BLAST-GENBANK

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
9	1104	T269 T816 T26 S27 T32 S4 S12 S123 S132 T182 S288 S342 S427 S453 S470 S506 T566 S626 T637 T752 T773 T826 S918 S933 S950 T959 S963 S101 T116 S123 S469 S686 S758 T769 T789 S831 S997 T1022 S1027 S1058	N212 N273 N1062	Transmembrane motif: Y69-R92 ATP-GTP binding site: G667-S674 C2 profile: L649-K735, I331-K417, L988-Q1077, L800-S877, L480-T558 C2 domain: V670-L682, E694-I707, L1014-E1039, L716-D724 G protein-coupled receptor: S881-L897 C2 domain: G973-D1097	g4193489 GLUT4 vesicle protein (Rattus norvegicus) Hashiramoto,M. (1998) Adv. Exp. Med. Biol. 441:47-61; Morris,N.J. et al. (1999) Biochim. Biophys.. Acta 1431:525-530	Motifs BLAST-GENBANK HMMER HMMER_PFAM BLIMPS_PRINTS BLAST_DOMO BLIMPS_PFAM
10	411	S15 S19 S29 S49 T84 S135 T209 T322 S387 T406 S19 S75 T117 T128 T322	N27 N126 N237		g3560143 putative vacuolar protein sorting-associated protein (Schizosaccharomyces pombe)	Motifs BLAST-GENBANK
11	201	S28 S141 S183 T44 Y171			g4689262 sorting nexin 11 (Homo sapiens)	Motifs BLAST-GENBANK

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
12	476	S61 S6 T44 S74 S156 S207 T245 T306 T330 T432 S6 T162 T245 S323	N43 N192 N305 N345 N385 N418		g3483017 TOM1-like protein (Homo sapiens) Seroussi,E. et al. (1999) Genomics 57:380-388	Motifs BLAST-GENBANK
13	1220	T1108 S81 T194 S233 S269 T434 S440 S465 T467 S527 S584 T674 S716 T1121 T1122 S1133 S66 S190 S211 T613 S799 S986 T1104 T1108 T1139 S1147 Y495 Y154 Y182	N40 N143 N562 N1181 N1197	WEB1 protein transport protein: M1-M229	g4104321 vesicle associated protein (Rattus norvegicus)	Motifs BLAST-GENBANK BLIMPS_PRODOM
14	222	T58 S93 T103	N197	Pleckstrin homology domain: A2-T109	g6013425 evectin-2 (Mus musculus) Krappa,R. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96:4633-4638	Motifs BLAST-GENBANK HMMER_PFAM

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
15	924	S431 S14 S145 T178 T211 T330 T349 S387 S435 S525 T563 T574 T594 T651 T675 S691 S698 S751 S889 T25 T77 S89 T118 S128 T231 T239 S269 S426 T440 T554 S631 S674 S876 Y534	N164 N174 N177 N411 N475	ATP-GTP binding site: A309-T316 Exocyst complex component: N494-S918	g2827158 rsec5 exocyst complex subunit (Rattus norvegicus) Kee,Y. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:14438-14443	Motifs BLAST-GENBANK BLAST_PRODOM
16	435	T48 T63 S239 S241 T256 S261 S308 S313 S370 S390 T409 T426 T44 S86 T137 T164 S317 S325 T398 Y230	N254 N404 N407 N418		g3560561 PAM COOH-terminal interactor protein (Rattus norvegicus) Chen,L. et al. (1998) J. Biol. Chem. 273:33524-33532	Motifs BLAST-GENBANK

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
17	321	T72 S130 S168 S226 S253 S280 S104 T264	N41 N262	Vacuolar sorting-associated protein: L17-P152	g3319953 TOM1 (Homo sapiens) Seroussi,E. et al. (1999) Genomics 57:380-388	Motifs BLAST-GENBANK BLAST_PRODOM
18	499	S89 S140 T207 S209 S223 T259 S267 S456 S463 S38 S72 T264 S351 T481	N75 N461	Signal peptide: M1-T25 Lectin precursor: S38-P235 Lumenal domain: P9-D215	g606828 er-Golgi mannose-specific lectin (Homo sapiens) Arar,C. et al. (1995) J. Biol. Chem. 270:3551-3553	Motifs BLAST-GENBANK BLAST_PRODOM BLAST_DOMO HMMER
19	879	T14 S104 T150 S183 S236 S237 S244 S271 T436 S476 S484 S486 T494 S505 T533 S559 T708 S731 S834 T864 T44 S58 S210 S297 S382 S440 S449 T627 S649 S705 Y403 Y687 Y785 Y814	N131 N146 N482 N703		g2078441 similar to S. cerevisiae intracellular protein transport protein (Caenorhabditis elegans) Wilson,R. et al. (1994) Nature 368:32-38	Motifs BLAST-GENBANK

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
20	298	T36 S77 S84 T126 S201 S270	N57 N158	NSF attachment protein signature: R37-K56, A100-F117, E129-E146, N162-Y185, M193-C212, P233-E253, K264-K284 NSF attachment protein: K19-D294 Soluble attachment protein SNAP: M1-E292	g3929617 alpha SNAP (platelet SNARE-associated protein) (Homo sapiens) Lemons,P.P. et al. (1997) Blood 90:1490-1500	Motifs BLAST-GENBANK BLIMPS_PRINTS BLAST_PRODOM BLAST_DOMO
21	941	S293 S478 T941 T50 S101 T147 S217 T299 T325 T359 S468 S478 S579 T625 S707 S718 S827 S849 S851 S853 S923 S54 S217 S306 S467 S656 T800 S813 T918 S937 Y532 Y737	N435 N763 N892	ATP/GTP binding site: A193-T200 Zinc finger C3HC4: C822-C860, K815-C860 PHD finger: H835-S849 ATP-binding vacuolar biogenesis protein: L146-D466	g666102 vacuolar biogenesis protein END1 (PEP5) protein (Saccharomyces cerevisiae) Woolford,C. (1990) Genetics 125:739-752	Motifs BLAST-GENBANK HMMER_PFAM BLIMPS_PFAM BLAST_PRODOM BLAST_DOMO

Table 2 cont.

Poly-peptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequence	Analytical Methods and Databases
22	336	T87 S130 T161 S43 T192 T267 T294 S301 T302 T318 T113 T125 T226 S227 T288	N117 N246	C2 domain: L47-E136, V196-R283, S34-V90 C2 domain signature: K63-L75, K92-K105, L137-D145 C2 Domain: L28-L137	g5926736 granuphilin-a (Mus musculus) Wang,J. et al. (1999) J. Biol. Chem. 274:28542-28548	Motifs BLAST-GENBANK HMMER_PFAM BLIMPS_PRINTS BLAST_DOMO PROFILESCAN
23	163	T11 T54 S145 T155 S86	N125	Ankyrin repeats: K64-R96, E97-V129, F130-K162	g6624055 similar to ankyrin motif (Homo sapiens) Sulston,J.E. and Waterston,R. (1998) Genome Res. 8:1097-1108	MOTIFS HMMER_PFAM

Table 3

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
24	1-292 686-1285 1715-1983	Gastrointestinal (0.250) Nervous (0.167) Cardiovascular (0.125) Reproductive (0. 125)	Cancer (0.417) Inflammation (0.208) Cell proliferation (0.167)	PBLUESCRIPT
25	1302-1559	Nervous (1.000)	Cancer (0.250) Inflammation (0.250) Trauma (0.167)	PBLUESCRIPT
26	1-386 1034-1179	Reproductive (0.538) Musculoskeletal (0.154)	Cancer (0.615) Inflammation (0.154) Cell proliferation (0.077)	PSPORT1
27	1-1124	Reproductive (0.255) Cardiovascular (0.149) Nervous (0.128)	Cancer (0.404) Cell proliferation (0.298) Inflammation (0.234)	pINCY
28	1-839 1697-2291	Reproductive (0.276) Gastrointestinal (0.161) Nervous (0.149)	Cancer (0.540) Inflammation (0.207) Cell proliferation (0.138)	pINCY
29	1-1509	Reproductive (0.213) Nervous (0.170) Endocrine (0.128) Gastrointestinal (0.128)	Cancer (0.447) Cell proliferation (0.191) Inflammation (0.191)	pINCY
30	659-1352 1976-2599	Reproductive (0.485) Cardiovascular (0.121) Gastrointestinal (0.121)	Cancer (0.576) Inflammation (0.212) Trauma (0.121)	pINCY
31	1-415	Reproductive (0.273) Nervous (0.227) Gastrointestinal (0.159)	Cancer (0.500) Cell proliferation (0.182) Inflammation (0.182)	PSPORT1
32	1-578 1020-3418	Reproductive (0.254) Hematopoietic/Immune (0.176) Nervous (0.162)	Cancer (0.401) Inflammation (0.310) Cell proliferation (0.176)	PBLUESCRIPT

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
33	696-916 1980-2046 2794-3343	Reproductive (0.255) Cardiovascular (0.157) Nervous (0.137)	Cancer (0.510) Inflammation (0.196) Trauma (0.157)	PSPORT1
34	473-730 1404-1496	Nervous (0.278) Hematopoietic/Immune (0.208) Gastrointestinal (0.125)	Cancer (0.417) Inflammation (0.347) Cell proliferation (0.153)	pINCY
35	1-42 1247-1482	Reproductive (0.341) Gastrointestinal (0.205) Urologic (0.114)	Cancer (0.636) Cell proliferation (0.114) Inflammation (0.114)	pINCY
36	1-948 1687-1972 2448-2534 2800-2960 3035-3074	Reproductive (0.231) Nervous (0.166) Hematopoietic/Immune (0.126)	Cancer (0.467) Inflammation (0.226) Cell proliferation (0.196)	pINCY
37	1-38	Hematopoietic/Immune (0.222) Nervous (0.204) Reproductive (0.204)	Inflammation (0.370) Cancer (0.352) Cell proliferation (0.259)	pINCY
38	1-71 565-625 1142-2946 3445-3899	Reproductive (0.250) Nervous (0.235) Gastrointestinal (0.132)	Cancer (0.368) Inflammation (0.265) Cell proliferation (0.162)	pINCY
39	1-33 601-723 993-1319 1467-1551	Cardiovascular (0.250) Dermatologic (0.250) Reproductive (0.250) Urologic (0.250)	Cancer (0.250) Inflammation (0.250) Trauma (0.250)	pINCY
40	1-72 651-1014 1066-1088	Reproductive (0.267) Gastrointestinal (0.200) Hematopoietic/Immune (0.133) Nervous (0.133) Urologic (0.133)	Cancer (0.333) Inflammation (0.267) Trauma (0.200)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
41	1-1367	Reproductive (0.75)	Cancer (0.75) Trauma (0.167)	pINCY
42	1-569 1069-2802	Nervous (0.193) Reproductive (0.193) Gastrointestinal (0.169)	Cancer (0.422) Cell proliferation (0.253) Inflammation (0.229)	pINCY
43		Nervous (0.600) Musculoskeletal (0.300) Reproductive (0.100)	Cancer (0.600) Neurological (0.200) Trauma (0.100)	pINCY
44	1-600	Hematopoietic/Immune (0.192) Reproductive (0.192) Nervous (0.179)	Cancer (0.410) Inflammation (0.295) Cell proliferation (0.167)	pINCY
45	1-1363	Reproductive (0.298) Gastrointestinal (0.234) Nervous (0.149)	Cancer (0.532) Inflammation (0.234) Cell proliferation (0.149)	pINCY
46	1-448 388-985 529-1034	Breast T-lymphocyte Prostate	Adenocarcinoma Ductal Type	

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
24	HYPONOB01	This library was constructed using RNA (Clontech, #6579-2) isolated from the hypothalamus tissues of 51 male and female Caucasian donors, 16 to 75 years old.
25	HYPONOB01	This library was constructed using RNA (Clontech, #6579-2) isolated from the hypothalamus tissues of 51 male and female Caucasian donors, 16 to 75 years old.
26	PGANNOT01	This library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male. Pathology indicated a benign paraganglioma and was associated with renal cell carcinoma, clear cell type, which did not penetrate the capsule.
27	LUNGFET03	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
28	LUNGNOT15	This library was constructed using RNA isolated from lung tissue removed from a 69-year-old Caucasian male. Pathology for the associated tumor tissue indicated residual invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, and malignant skin neoplasm. Family history included cerebrovascular disease, type I diabetes, acute myocardial infarction, and arteriosclerotic coronary disease.
29	BRAITUT13	This library was constructed using RNA isolated from brain tumor tissue from the frontal lobe of a 68-year-old Caucasian male. Pathology indicated a meningioma in the frontal lobe.
30	LIVRTUT01	This library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female. Pathology indicated metastatic adenocarcinoma consistent with colon cancer. Family history included malignant neoplasm of the liver.
31	BRAINON01	This library was constructed and normalized from 4.88 million independent clones from a brain library. RNA was made from brain tissue from a 26-year-old Caucasian male. Pathology for the associated tumor tissue indicated oligoastrocytoma in the right fronto-parietal part of the brain.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
32	UCMCL5T01	This library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
33	BRAINON01	This library was constructed and normalized from 4.88 million independent clones from a brain library. RNA was made from brain tissue from a 26-year-old Caucasian male. Pathology for the associated tumor tissue indicated oligoastrocytoma in the right fronto-parietal part of the brain.
34	ISLTNOT01	This library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
35	SMCANOT01	This library was constructed using RNA isolated from an aortic smooth muscle cell line derived from the explanted heart of a male during a heart transplant.
36	CONUTUT01	This library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
37	OVARTUT02	This library was constructed using RNA isolated from ovarian tumor tissue removed from a 51-year-old Caucasian female during an exploratory laparotomy, total abdominal hysterectomy, salpingo-oophorectomy. Pathology indicated mucinous cystadenoma. Family history included atherosclerotic coronary artery disease, benign hypertension, breast cancer, and uterine cancer.
38	LUNGNOT23	This library was constructed using RNA isolated from lung tissue from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
39	BRSTNOT12	This library was constructed using RNA isolated from diseased breast tissue removed from a 32-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included benign hypertension and atherosclerotic coronary artery disease.
40	UTRSNON03	This normalized library was constructed from 6.4 million independent clones from a uterus library. RNA was isolated from uterine myometrial tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology for the associated tumor tissue indicated uterine leiomyoma. Patient history included ventral hernia and a benign ovarian neoplasm. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228).
41	PROSTMT03	This library was constructed using RNA isolated from prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy and regional lymph node excision. Pathology for the associated tumor indicated adenocarcinoma. The patient presented with elevated prostate specific antigen (PSA) and induration. Patient history included pure hypercholesterolemia, kidney calculus, an unspecified allergy, and atopic dermatitis. Family history included colon cancer.
42	SINTNOT25	This library was constructed using RNA isolated from small intestine tissue removed from a 13-year-old Caucasian male, who died from a gunshot wound to the head. Family history included diabetes.
43	BRSTDIT01	This library was constructed using RNA isolated from diseased breast tissue from a 48-year-old Caucasian female. Pathology for the associated tumor tissue indicated intraductal cancer. The patient presented with a malignant neoplasm of the breast and unspecified breast symptoms. Patient history included mitral valve disorder and an unspecified disease of the shoulder region. Family history included malignant neoplasm of the breast and hyperlipidemia, malignant neoplasm of the colon and cardiac dysrhythmias, and malignant neoplasm of the colon.
44	BRAENOT02	This library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
45	COLTDIT04	This library was constructed from diseased transverse colon tissue removed from a 16-year-old Caucasian male during partial colectomy, temporary ileostomy, and colonoscopy. Pathology indicated innumerable (greater than 100) adenomatous polyps with low-grade dysplasia involving the entire colonic mucosa in the setting of familial polyposis coli. The anal mucosa showed 10 adenomatous polyps with low-grade dysplasia in the setting of familial polyposis coli. The patient presented with abdominal pain and flatulence. Family history included benign colon neoplasm in the father; benign colon neoplasm in the sibling(s); and benign hypertension, cerebrovascular disease, breast cancer, uterine cancer, and type II diabetes in the grandparent(s).

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-23,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-23.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:24-46.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

18. A method for treating a disease or condition associated with decreased expression of functional VETRP, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional VETRP, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional VETRP, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
TANG, Y. Tom
YUE, Henry
BANDMAN, Olga
HILLMAN, Jennifer
BAUGHN, Mariah R.
LU, Dyung Aina M.
AZIMZAI, Yalda

<120> VESICLE TRAFFICKING PROTEINS

<130> PF-0756 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/172,968; 60/172,066
<151> 1999-12-21; 1999-12-23

<160> 46
<170> PERL Program

<210> 1
<211> 481
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 381039CD1

<400> 1
Met Val Met Ala Ala Lys Lys Gly Pro Gly Pro Gly Gly Gly Val
1 5 10 15
Ser Gly Gly Lys Ala Glu Ala Glu Ala Ala Ser Glu Val Trp Cys
20 25 30
Arg Arg Val Arg Glu Leu Gly Gly Cys Ser Gln Ala Gly Asn Arg
35 40 45
His Cys Phe Glu Cys Ala Gln Arg Gly Val Thr Tyr Val Asp Ile
50 55 60
Thr Val Gly Ser Phe Val Cys Thr Thr Cys Ser Gly Leu Leu Arg
65 70 75
Gly Leu Asn Pro Pro His Arg Val Lys Ser Ile Ser Met Thr Thr
80 85 90
Phe Thr Glu Pro Glu Val Val Phe Leu Gln Ser Arg Gly Asn Glu
95 100 105
Val Cys Arg Lys Ile Trp Leu Gly Leu Phe Asp Ala Arg Thr Ser
110 115 120
Leu Val Pro Asp Ser Arg Asp Pro Gln Lys Val Lys Glu Phe Leu
125 130 135
Gln Glu Lys Tyr Glu Lys Lys Arg Trp Tyr Val Pro Pro Asp Gln
140 145 150
Val Lys Gly Pro Thr Tyr Thr Lys Gly Ser Ala Ser Thr Pro Val
155 160 165
Gln Gly Ser Ile Pro Glu Gly Lys Pro Leu Arg Thr Leu Leu Gly
170 175 180
Asp Pro Ala Pro Ser Leu Ser Val Ala Ala Ser Thr Ser Ser Gln
185 190 195
Pro Val Ser Gln Ser His Ala Arg Thr Ser Gln Ala Arg Ser Thr
200 205 210
Gln Pro Pro Pro His Ser Ser Val Lys Lys Ala Ser Thr Asp Leu
215 220 225
Leu Ala Asp Ile Gly Gly Asp Pro Phe Ala Ala Pro Gln Met Ala
230 235 240
Pro Ala Phe Ala Ala Phe Pro Ala Phe Gly Gly Gln Thr Pro Ser
245 250 255
Gln Gly Gly Phe Ala Asn Phe Asp Ala Phe Ser Ser Gly Pro Ser

260 265 270
Ser Ser Val Phe Gly Ser Leu Pro Pro Ala Gly Gln Ala Ser Phe
275 280 285
Gln Ala Gln Pro Thr Pro Ala Gly Ser Ser Gln Gly Thr Pro Phe
290 295 300
Gly Ala Thr Pro Leu Ala Pro Ala Ser Gln Pro Asn Ser Leu Ala
305 310 315
Asp Val Gly Ser Phe Leu Gly Pro Gly Val Pro Ala Ala Gly Val
320 325 330
Pro Ser Ser Leu Phe Gly Met Ala Gly Gln Val Pro Pro Leu Gln
335 340 345
Ser Val Thr Met Gly Gly Gly Gly Gly Ser Ser Thr Gly Leu Ala
350 355 360
Phe Gly Ala Phe Thr Asn Pro Phe Thr Ala Pro Ala Ala Gln Ser
365 370 375
Pro Leu Pro Ser Thr Asn Pro Phe Gln Pro Asn Gly Leu Ala Pro
380 385 390
Gly Pro Gly Phe Gly Met Ser Ser Ala Gly Pro Gly Phe Pro Gln
395 400 405
Ala Val Pro Pro Thr Gly Ala Phe Ala Ser Ser Phe Pro Ala Pro
410 415 420
Leu Phe Pro Pro Gln Thr Pro Leu Val Gln Gln Gln Asn Gly Ser
425 430 435
Ser Phe Gly Asp Leu Gly Ser Ala Lys Leu Gly Gln Arg Pro Leu
440 445 450
Ser Gln Pro Ala Gly Ile Ser Thr Asn Pro Phe Met Thr Gly Pro
455 460 465
Ser Ser Ser Pro Phe Ala Ser Lys Pro Pro Thr Thr Asn Pro Phe
470 475 480
Leu

<210> 2
<211> 195
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 383249CD1

<400> 2
Met Ser Ser Cys Ser Asn Val Cys Gly Ser Arg Gln Ala Gln Ala
1 5 10 15
Ala Ala Glu Gly Gly Tyr Gln Arg Tyr Gly Val Arg Ser Tyr Leu
20 25 30
His Gln Phe Tyr Glu Asp Cys Thr Ala Ser Ile Trp Glu Tyr Glu
35 40 45
Asp Asp Phe Gln Ile Gln Arg Ser Pro Asn Arg Trp Ser Ser Val
50 55 60
Phe Trp Lys Val Gly Leu Ile Ser Gly Thr Val Phe Val Ile Leu
65 70 75
Gly Leu Thr Val Leu Ala Val Gly Phe Leu Val Pro Pro Lys Ile
80 85 90
Glu Ala Phe Gly Glu Ala Asp Phe Val Val Val Asp Thr His Ala
95 100 105
Val Gln Phe Asn Ser Ala Leu Asp Met Tyr Lys Leu Ala Gly Ala
110 115 120
Val Leu Phe Cys Ile Gly Gly Thr Ser Met Ala Gly Cys Leu Leu
125 130 135
Met Ser Val Phe Val Lys Ser Tyr Ser Lys Glu Glu Lys Phe Leu
140 145 150
Gln Gln Lys Phe Lys Glu Arg Ile Ala Asp Ile Lys Ala His Thr
155 160 165
Gln Pro Val Thr Lys Ala Pro Gly Pro Gly Glu Thr Lys Ile Pro
170 175 180
Val Thr Leu Ser Arg Val Gln Asn Val Gln Pro Leu Leu Ala Thr

185 190 195

<210> 3
<211> 313
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 618769CD1

<400> 3
Met Gly Asn Phe Asp Asn Ala Asn Val Thr Gly Glu Ile Glu Phe
1 5 10 15
Ala Ile His Tyr Cys Phe Lys Thr His Ser Leu Glu Ile Cys Ile
20 25 30
Lys Ala Cys Lys Asn Leu Ala Tyr Gly Glu Glu Lys Lys Lys Lys
35 40 45
Cys Asn Pro Tyr Val Lys Thr Tyr Leu Leu Pro Asp Arg Ser Ser
50 55 60
Gln Gly Lys Arg Lys Thr Gly Val Gln Arg Asn Thr Val Asp Pro
65 70 75
Thr Phe Gln Glu Thr Leu Lys Tyr Gln Val Ala Pro Ala Gln Leu
80 85 90
Val Thr Arg Gln Leu Gln Val Ser Val Trp His Leu Gly Thr Leu
95 100 105
Ala Arg Arg Val Phe Leu Gly Glu Val Ile Ile Pro Leu Ala Thr
110 115 120
Trp Asp Phe Glu Asp Ser Thr Thr Gln Ser Phe Arg Trp His Pro
125 130 135
Leu Arg Ala Lys Ala Glu Lys Tyr Glu Asp Ser Val Pro Gln Ser
140 145 150
Asn Gly Glu Leu Thr Val Arg Ala Lys Leu Val Leu Pro Ser Arg
155 160 165
Pro Arg Lys Leu Gln Glu Ala Gln Glu Gly Thr Asp Gln Pro Ser
170 175 180
Leu His Gly Gln Leu Cys Leu Val Val Leu Gly Ala Lys Asn Leu
185 190 195
Pro Val Arg Pro Asp Gly Thr Leu Asn Ser Phe Val Lys Gly Cys
200 205 210
Leu Thr Leu Pro Asp Gln Gln Lys Leu Arg Leu Lys Ser Pro Val
215 220 225
Leu Arg Lys Gln Ala Cys Pro Gln Trp Lys His Ser Phe Val Phe
230 235 240
Ser Gly Val Thr Pro Ala Gln Leu Arg Gln Ser Ser Leu Glu Leu
245 250 255
Thr Val Trp Asp Gln Ala Leu Phe Gly Met Asn Asp Arg Leu Leu
260 265 270
Gly Gly Thr Arg Leu Gly Ser Lys Gly Asp Thr Ala Val Gly Gly
275 280 285
Asp Ala Cys Ser Leu Ser Lys Leu Gln Trp Gln Lys Val Leu Ser
290 295 300
Ser Pro Asn Leu Trp Thr Asp Met Thr Leu Val Leu His
305 310

<210> 4
<211> 201
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1234837CD1

<400> 4
Met Gly Asn Leu Phe Gly Arg Lys Lys Gln Ser Arg Val Thr Glu

```
 1               5                  10                  15
Gln Asp Lys Ala Ile Leu Gln Leu Lys Gln Gln Arg Asp Lys Leu
            20                  25                  30
Arg Gln Tyr Gln Lys Arg Ile Ala Gln Gln Leu Glu Arg Glu Arg
        35                  40                  45
Ala Leu Ala Arg Gln Leu Leu Arg Asp Gly Arg Lys Glu Arg Ala
    50                  55                  60
Lys Leu Leu Leu Lys Lys Lys Arg Tyr Gln Glu Gln Leu Leu Asp
65                  70                  75
Arg Thr Glu Asn Gln Ile Ser Ser Leu Glu Ala Met Val Gln Ser
                80                  85                  90
Ile Glu Phe Thr Gln Ile Glu Met Lys Val Met Glu Gly Leu Gln
            95                  100                 105
Phe Gly Asn Glu Cys Leu Asn Lys Met His Gln Val Met Ser Ile
        110                 115                 120
Glu Glu Val Glu Arg Ile Leu Asp Glu Thr Gln Glu Ala Val Glu
    125                 130                 135
Tyr Gln Arg Gln Ile Asp Glu Leu Leu Ala Gly Ser Phe Thr Gln
140                 145                 150
Glu Asp Glu Asp Ala Ile Leu Glu Glu Leu Ser Ala Ile Thr Gln
155                 160                 165
Glu Gln Ile Glu Leu Pro Glu Val Pro Ser Glu Pro Leu Pro Glu
                170                 175                 180
Lys Ile Pro Glu Asn Val Pro Val Lys Ala Arg Pro Arg Gln Ala
            185                 190                 195
Glu Leu Val Ala Ala Ser
        200
```

<210> 5
<211> 566
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1607223CD1

<400> 5

```
Met Lys Ser Cys Gly Lys Arg Phe His Asp Glu Val Gly Lys Phe
 1               5                  10                  15
Arg Phe Leu Asn Glu Leu Ile Lys Val Val Ser Pro Lys Tyr Leu
            20                  25                  30
Gly Ser Arg Thr Ser Glu Lys Val Lys Asn Lys Ile Leu Glu Leu
        35                  40                  45
Leu Tyr Ser Trp Thr Val Gly Leu Pro Glu Glu Val Lys Ile Ala
    50                  55                  60
Glu Ala Tyr Gln Met Leu Lys Lys Gln Gly Ile Val Lys Ser Asp
65                  70                  75
Pro Lys Leu Pro Asp Asp Thr Thr Phe Pro Leu Pro Pro Pro Arg
                80                  85                  90
Pro Lys Asn Val Ile Phe Glu Asp Glu Glu Lys Ser Lys Met Leu
            95                  100                 105
Ala Arg Leu Leu Lys Ser Ser His Pro Glu Asp Leu Arg Ala Ala
        110                 115                 120
Asn Lys Leu Ile Lys Glu Met Val Gln Glu Asp Gln Lys Arg Met
    125                 130                 135
Glu Lys Ile Ser Lys Arg Val Asn Ala Ile Glu Glu Val Asn Asn
140                 145                 150
Asn Val Lys Leu Leu Thr Glu Met Val Met Ser His Ser Gln Gly
155                 160                 165
Gly Ala Ala Ala Gly Ser Ser Glu Asp Leu Met Lys Glu Leu Tyr
                170                 175                 180
Gln Arg Cys Glu Arg Met Arg Pro Thr Leu Phe Arg Leu Ala Ser
            185                 190                 195
Asp Thr Glu Asp Asn Asp Glu Ala Leu Ala Glu Ile Leu Gln Ala
        200                 205                 210
Asn Asp Asn Leu Thr Gln Val Ile Asn Leu Tyr Lys Gln Leu Val
```

215 220 225
Arg Gly Glu Glu Val Asn Gly Asp Ala Thr Ala Gly Ser Ile Pro
230 235 240
Gly Ser Thr Ser Ala Leu Leu Asp Leu Ser Gly Leu Asp Leu Pro
245 250 255
Pro Ala Gly Thr Thr Tyr Pro Ala Met Pro Thr Arg Pro Gly Glu
260 265 270
Gln Ala Ser Pro Glu Gln Pro Ser Ala Ser Val Ser Leu Leu Asp
275 280 285
Asp Glu Leu Met Ser Leu Gly Leu Ser Asp Pro Thr Pro Pro Ser
290 295 300
Gly Pro Ser Leu Asp Gly Thr Gly Trp Asn Ser Phe Gln Ser Ser
305 310 315
Asp Ala Thr Glu Pro Pro Ala Pro Ala Leu Ala Gln Ala Pro Ser
320 325 330
Met Glu Ser Arg Pro Pro Ala Gln Thr Ser Leu Pro Ala Ser Ser
335 340 345
Gly Leu Asp Asp Leu Asp Leu Leu Gly Lys Thr Leu Leu Gln Gln
350 355 360
Ser Leu Pro Pro Glu Ser Gln Gln Val Arg Trp Glu Lys Gln Gln
365 370 375
Pro Thr Pro Arg Leu Thr Leu Arg Asp Leu Gln Asn Lys Ser Ser
380 385 390
Ser Cys Ser Ser Pro Ser Ser Ser Ala Thr Ser Leu Leu His Thr
395 400 405
Val Ser Pro Glu Pro Pro Arg Pro Pro Gln Gln Pro Val Pro Thr
410 415 420
Glu Leu Ser Leu Ala Ser Ile Thr Val Pro Leu Glu Ser Ile Lys
425 430 435
Pro Ser Asn Ile Leu Pro Val Thr Val Tyr Asp Gln His Gly Phe
440 445 450
Arg Ile Leu Phe His Phe Ala Arg Asp Pro Leu Pro Gly Arg Ser
455 460 465
Asp Val Leu Val Val Val Ser Met Leu Ser Thr Ala Pro Gln
470 475 480
Pro Ile Arg Asn Ile Val Phe Gln Ser Ala Val Pro Lys Val Met
485 490 495
Lys Val Lys Leu Gln Pro Pro Ser Gly Thr Glu Leu Pro Ala Phe
500 505 510
Asn Pro Ile Val His Pro Ser Ala Ile Thr Gln Val Leu Leu Leu
515 520 525
Ala Asn Pro Gln Lys Glu Lys Val Arg Leu Arg Tyr Lys Leu Thr
530 535 540
Phe Thr Met Gly Asp Gln Thr Tyr Asn Glu Met Gly Asp Val Asp
545 550 555
Gln Phe Pro Pro Pro Glu Thr Trp Gly Ser Leu
560 565

<210> 6
<211> 270
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1621554CD1

<400> 6
Met Gly Phe Trp Cys Arg Met Ser Glu Asn Gln Glu Gln Glu Glu
1 5 10 15
Val Ile Thr Val Arg Val Gln Asp Pro Arg Val Gln Asn Glu Gly
20 25 30
Ser Trp Asn Ser Tyr Val Asp Tyr Lys Ile Phe Leu His Thr Asn
35 40 45
Ser Lys Ala Phe Thr Ala Lys Thr Ser Cys Val Arg Arg Arg Tyr
50 55 60
Arg Glu Phe Val Trp Leu Arg Lys Gln Leu Gln Arg Asn Ala Gly

65 70 75
Leu Val Pro Val Pro Glu Leu Pro Gly Lys Ser Thr Phe Phe Gly
80 85 90
Thr Ser Asp Glu Phe Ile Glu Lys Arg Arg Gln Gly Leu Gln His
95 100 105
Phe Leu Glu Lys Val Leu Gln Ser Val Val Leu Leu Ser Asp Ser
110 115 120
Gln Leu His Leu Phe Leu Gln Ser Gln Leu Ser Val Pro Glu Ile
125 130 135
Glu Ala Cys Val Gln Gly Arg Ser Thr Met Thr Val Ser Asp Ala
140 145 150
Ile Leu Arg Tyr Ala Met Ser Asn Cys Gly Trp Ala Gln Glu Glu
155 160 165
Arg Gln Ser Ser Ser His Leu Ala Lys Gly Asp Gln Pro Lys Ser
170 175 180
Cys Cys Phe Leu Pro Arg Ser Gly Arg Arg Ser Ser Pro Ser Pro
185 190 195
Pro Pro Ser Glu Glu Lys Asp His Leu Glu Val Trp Ala Pro Val
200 205 210
Val Asp Ser Glu Val Pro Ser Leu Glu Ser Pro Thr Leu Pro Pro
215 220 225
Leu Ser Ser Pro Leu Cys Cys Asp Phe Gly Arg Pro Lys Glu Gly
230 235 240
Thr Ser Thr Leu Gln Ser Val Arg Arg Ala Val Gly Gly Asp His
245 250 255
Ala Val Pro Leu Asp Pro Gly Gln Leu Glu Thr Val Leu Glu Lys
260 265 270

<210> 7
<211> 490
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1751553CD1

<400> 7
Met Ala Thr Glu Phe Ile Lys Ser Cys Cys Gly Gly Cys Phe Tyr
1 5 10 15
Gly Glu Thr Glu Lys His Asn Phe Ser Val Glu Arg Asp Phe Lys
20 25 30
Ala Ala Val Pro Asn Ser Gln Asn Ala Thr Ile Ser Val Pro Pro
35 40 45
Leu Thr Ser Val Ser Val Lys Pro Gln Leu Gly Cys Thr Glu Asp
50 55 60
Tyr Leu Leu Ser Lys Leu Pro Ser Asp Gly Lys Glu Val Pro Phe
65 70 75
Val Val Pro Lys Phe Lys Leu Ser Tyr Ile Gln Pro Arg Thr Gln
80 85 90
Glu Thr Pro Ser His Leu Glu Glu Leu Glu Gly Ser Ala Arg Ala
95 100 105
Ser Phe Gly Asp Arg Lys Val Glu Leu Ser Ser Ser Ser Gln His
110 115 120
Gly Pro Ser Tyr Asp Val Tyr Asn Pro Phe Tyr Met Tyr Gln His
125 130 135
Ile Ser Pro Asp Leu Ser Arg Arg Phe Pro Pro Arg Ser Glu Val
140 145 150
Thr Arg Leu Tyr Gly Ser Val Cys Asp Leu Arg Thr Asn Lys Leu
155 160 165
Pro Gly Ser Pro Gly Leu Ser Lys Ser Met Phe Asp Leu Thr Asn
170 175 180
Ser Ser Gln Arg Phe Ile Gln Arg His Asp Ser Leu Ser Ser Val
185 190 195
Pro Ser Ser Ser Ser Ser Arg Lys Asn Ser Gln Gly Ser Asn Arg
200 205 210

Ser Leu Asp Thr Ile Thr Leu Ser Gly Asp Glu Arg Asp Phe Gly
215 220 225
Arg Leu Asn Val Lys Leu Phe Tyr Asn Ser Ser Val Glu Gln Ile
230 235 240
Trp Ile Thr Val Leu Gln Cys Arg Asp Leu Ser Trp Pro Ser Ser
245 250 255
Tyr Gly Asp Thr Pro Thr Val Ser Ile Lys Gly Ile Leu Thr Leu
260 265 270
Pro Lys Pro Val His Phe Lys Ser Ser Ala Lys Glu Gly Ser Asn
275 280 285
Ala Ile Glu Phe Met Glu Thr Phe Val Phe Ala Ile Lys Leu Gln
290 295 300
Asn Leu Gln Thr Val Arg Leu Val Phe Lys Ile Gln Thr Gln Thr
305 310 315
Pro Arg Lys Lys Thr Ile Gly Glu Cys Ser Met Ser Leu Arg Thr
320 325 330
Leu Ser Thr Gln Glu Met Asp Tyr Ser Leu Asp Ile Thr Pro Pro
335 340 345
Ser Lys Ile Ser Val Cys His Ala Glu Leu Glu Leu Gly Thr Cys
350 355 360
Phe Gln Ala Val Asn Ser Arg Ile Gln Leu Gln Ile Leu Glu Ala
365 370 375
Arg Tyr Leu Pro Ser Ser Ser Thr Pro Leu Thr Leu Ser Phe Phe
380 385 390
Val Lys Val Gly Met Phe Ser Ser Gly Glu Leu Ile Tyr Lys Lys
395 400 405
Lys Thr Arg Leu Leu Lys Ala Ser Asn Gly Arg Val Lys Trp Gly
410 415 420
Glu Thr Met Ile Phe Pro Leu Ile Gln Ser Glu Lys Glu Ile Val
425 430 435
Phe Leu Ile Lys Leu Tyr Ser Arg Ser Ser Val Arg Arg Lys His
440 445 450
Phe Val Gly Gln Ile Trp Ile Ser Glu Asp Ser Asn Asn Ile Glu
455 460 465
Ala Val Asn Gln Trp Lys Glu Thr Val Ile Asn Pro Glu Lys Val
470 475 480
Val Ile Arg Trp His Lys Leu Asn Pro Ser
485 490

<210> 8
<211> 136
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1832403CD1

<400> 8
Met Ala Gly Ala Gly Ser Ala Ala Val Ser Gly Ala Gly Thr Pro
1 5 10 15
Val Ala Gly Pro Thr Gly Arg Asp Leu Phe Ala Glu Gly Leu Leu
20 25 30
Glu Phe Leu Arg Pro Ala Val Gln Gln Leu Asp Ser His Val His
35 40 45
Ala Val Arg Glu Ser Gln Val Glu Leu Arg Glu Gln Ile Asp Asn
50 55 60
Leu Ala Thr Glu Leu Cys Arg Ile Asn Glu Asp Gln Lys Val Ala
65 70 75
Leu Asp Leu Asp Pro Tyr Val Lys Lys Leu Leu Asn Ala Arg Arg
80 85 90
Arg Val Val Leu Val Asn Asn Ile Leu Gln Asn Ala Gln Glu Arg
95 100 105
Leu Arg Arg Leu Asn His Ser Val Ala Lys Glu Thr Ala Arg Arg
110 115 120
Arg Ala Met Leu Asp Ser Gly Ile Tyr Pro Pro Gly Ser Pro Gly
125 130 135

Lys

<210> 9
<211> 1104
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1971747CD1

<400> 9
Met Glu Arg Ser Pro Gly Glu Gly Pro Ser Pro Ser Pro Met Asp
1 5 10 15
Gln Pro Ser Ala Pro Ser Asp Pro Thr Asp Gln Pro Pro Ala Ala
20 25 30
His Ala Lys Pro Asp Pro Gly Ser Gly Gly Gln Pro Ala Gly Pro
35 40 45
Gly Ala Ala Gly Glu Ala Leu Ala Val Leu Thr Ser Phe Gly Arg
50 55 60
Arg Leu Leu Val Leu Ile Pro Val Tyr Leu Ala Gly Ala Val Gly
65 70 75
Leu Ser Val Gly Phe Val Leu Phe Gly Leu Ala Leu Tyr Leu Gly
80 85 90
Trp Arg Arg Val Arg Asp Glu Lys Glu Arg Ser Leu Arg Ala Ala
95 100 105
Arg Gln Leu Leu Asp Asp Glu Glu Gln Leu Thr Ala Lys Thr Leu
110 115 120
Tyr Met Ser His Arg Glu Leu Pro Ala Trp Val Ser Phe Pro Asp
125 130 135
Val Glu Lys Ala Glu Trp Leu Asn Lys Ile Val Ala Gln Val Trp
140 145 150
Pro Phe Leu Gly Gln Tyr Met Glu Lys Leu Leu Ala Glu Thr Val
155 160 165
Ala Pro Ala Val Arg Gly Ser Asn Pro His Leu Gln Thr Phe Thr
170 175 180
Phe Thr Arg Val Glu Leu Gly Glu Lys Pro Leu Arg Ile Ile Gly
185 190 195
Val Lys Val His Pro Gly Gln Arg Lys Glu Gln Ile Leu Leu Asp
200 205 210
Leu Asn Ile Ser Tyr Val Gly Asp Val Gln Ile Asp Val Glu Val
215 220 225
Lys Lys Tyr Phe Cys Lys Ala Gly Val Lys Gly Met Gln Leu His
230 235 240
Gly Val Leu Arg Val Ile Leu Glu Pro Leu Ile Gly Asp Leu Pro
245 250 255
Phe Val Gly Ala Val Ser Met Phe Phe Ile Arg Arg Pro Thr Leu
260 265 270
Asp Ile Asn Trp Thr Gly Met Thr Asn Leu Leu Asp Ile Pro Gly
275 280 285
Leu Ser Ser Leu Ser Asp Thr Met Ile Met Asp Ser Ile Ala Ala
290 295 300
Phe Leu Val Leu Pro Asn Arg Leu Leu Val Pro Leu Val Pro Asp
305 310 315
Leu Gln Asp Val Ala Gln Leu Arg Ser Pro Leu Pro Arg Gly Ile
320 325 330
Ile Arg Ile His Leu Leu Ala Ala Arg Gly Leu Ser Ser Lys Asp
335 340 345
Lys Tyr Val Lys Gly Leu Ile Glu Gly Lys Ser Asp Pro Tyr Ala
350 355 360
Leu Val Arg Leu Gly Thr Gln Thr Phe Cys Ser Arg Val Ile Asp
365 370 375
Glu Glu Leu Asn Pro Gln Trp Gly Glu Thr Tyr Glu Val Met Val
380 385 390
His Glu Val Pro Gly Gln Glu Ile Glu Val Glu Val Phe Asp Lys
395 400 405

Asp Pro Asp Lys Asp Asp Phe Leu Gly Arg Met Lys Leu Asp Val
410 415 420
Gly Lys Val Leu Gln Ala Ser Val Leu Asp Asp Trp Phe Pro Leu
425 430 435
Gln Gly Gly Gln Gly Gln Val His Leu Arg Leu Glu Trp Leu Ser
440 445 450
Leu Leu Ser Asp Ala Glu Lys Leu Glu Gln Val Leu Gln Trp Asn
455 460 465
Trp Gly Val Ser Ser Arg Pro Asp Pro Pro Ser Ala Ala Ile Leu
470 475 480
Val Val Tyr Leu Asp Arg Ala Gln Asp Leu Pro Leu Lys Lys Gly
485 490 495
Asn Lys Glu Pro Asn Pro Met Val Gln Leu Ser Ile Gln Asp Val
500 505 510
Thr Gln Glu Ser Lys Ala Val Tyr Ser Thr Asn Cys Pro Val Trp
515 520 525
Glu Glu Ala Phe Arg Phe Phe Leu Gln Asp Pro Gln Ser Gln Glu
530 535 540
Leu Asp Val Gln Val Lys Asp Asp Ser Arg Ala Leu Thr Leu Gly
545 550 555
Ala Leu Thr Leu Pro Leu Ala Arg Leu Leu Thr Ala Pro Glu Leu
560 565 570
Ile Leu Asp Gln Trp Phe Gln Leu Ser Ser Ser Gly Pro Asn Ser
575 580 585
Arg Leu Tyr Met Lys Leu Val Met Arg Ile Leu Tyr Leu Asp Ser
590 595 600
Ser Glu Ile Cys Phe Pro Thr Val Pro Gly Cys Pro Gly Ala Trp
605 610 615
Asp Val Asp Ser Glu Asn Pro Gln Arg Gly Ser Ser Val Asp Ala
620 625 630
Pro Pro Arg Pro Cys His Thr Thr Pro Asp Ser Gln Phe Gly Thr
635 640 645
Glu His Val Leu Arg Ile His Val Leu Glu Ala Gln Asp Leu Ile
650 655 660
Ala Lys Asp Arg Phe Leu Gly Gly Leu Val Lys Gly Lys Ser Asp
665 670 675
Pro Tyr Val Lys Leu Lys Leu Ala Gly Arg Ser Phe Arg Ser His
680 685 690
Val Val Arg Glu Asp Leu Asn Pro Arg Trp Asn Glu Val Phe Glu
695 700 705
Val Ile Val Thr Ser Val Pro Gly Gln Glu Leu Glu Val Glu Val
710 715 720
Phe Asp Lys Asp Leu Asp Lys Asp Asp Phe Leu Gly Arg Cys Lys
725 730 735
Val Arg Leu Thr Thr Val Leu Asn Ser Gly Phe Leu Asp Glu Trp
740 745 750
Leu Thr Leu Glu Asp Val Pro Ser Gly Arg Leu His Leu Arg Leu
755 760 765
Glu Arg Leu Thr Pro Arg Pro Thr Ala Ala Glu Leu Glu Glu Val
770 775 780
Leu Gln Val Asn Ser Leu Ile Gln Thr Gln Lys Ser Ala Glu Leu
785 790 795
Ala Ala Ala Leu Leu Ser Ile Tyr Met Glu Arg Ala Glu Asp Leu
800 805 810
Pro Leu Arg Lys Gly Thr Lys His Leu Ser Pro Tyr Ala Thr Leu
815 820 825
Thr Val Gly Asp Ser Ser His Lys Thr Lys Thr Ile Ser Gln Thr
830 835 840
Ser Ala Pro Val Trp Asp Glu Ser Ala Ser Phe Leu Ile Arg Lys
845 850 855
Pro His Thr Glu Ser Leu Glu Leu Gln Val Arg Gly Glu Gly Thr
860 865 870
Gly Val Leu Gly Ser Leu Ser Leu Pro Leu Ser Glu Leu Leu Val
875 880 885
Ala Asp Gln Leu Cys Leu Asp Arg Trp Phe Thr Leu Ser Ser Gly
890 895 900
Gln Gly Gln Val Leu Leu Arg Ala Gln Leu Gly Ile Leu Val Ser

905 910 915
Gln His Ser Gly Val Glu Ala His Ser His Ser Tyr Ser His Ser
920 925 930
Ser Ser Ser Leu Ser Glu Glu Pro Glu Leu Ser Gly Gly Pro Pro
935 940 945
His Ile Thr Ser Ser Ala Pro Glu Leu Arg Gln Arg Leu Thr His
950 955 960
Val Asp Ser Pro Leu Glu Ala Pro Ala Gly Pro Leu Gly Gln Val
965 970 975
Lys Leu Thr Leu Trp Tyr Tyr Ser Glu Glu Arg Lys Leu Val Ser
980 985 990
Ile Val His Gly Cys Arg Ser Leu Arg Gln Asn Gly Arg Asp Pro
995 1000 1005
Pro Asp Pro Tyr Val Ser Leu Leu Leu Leu Pro Asp Lys Asn Arg
1010 1015 1020
Gly Thr Lys Arg Arg Thr Ser Gln Lys Lys Arg Thr Leu Ser Pro
1025 1030 1035
Glu Phe Asn Glu Arg Phe Glu Trp Glu Leu Pro Leu Asp Glu Ala
1040 1045 1050
Gln Arg Arg Lys Leu Asp Val Ser Val Lys Ser Asn Ser Ser Phe
1055 1060 1065
Met Ser Arg Glu Arg Glu Leu Leu Gly Lys Val Gln Leu Asp Leu
1070 1075 1080
Ala Glu Thr Asp Leu Ser Gln Gly Val Ala Arg Trp Tyr Asp Leu
1085 1090 1095
Met Asp Asn Lys Asp Lys Gly Ser Ser
1100

<210> 10
<211> 411
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2285348CD1

<400> 10
Met Ala Asn Tyr Glu Ser Thr Glu Val Met Gly Asp Gly Glu Ser
1 5 10 15
Ala His Asp Ser Pro Arg Asp Glu Ala Leu Gln Asn Ile Ser Ala
20 25 30
Asp Asp Leu Pro Asp Ser Ala Ser Gln Ala Ala His Pro Gln Asp
35 40 45
Ser Ala Phe Ser Tyr Arg Asp Ala Lys Lys Lys Leu Arg Leu Ala
50 55 60
Leu Cys Ser Ala Asp Ser Val Ala Phe Pro Val Leu Thr His Ser
65 70 75
Thr Arg Asn Gly Leu Pro Asp His Thr Asp Pro Glu Asp Asn Glu
80 85 90
Ile Val Cys Phe Leu Lys Val Gln Ile Ala Glu Ala Ile Asn Leu
95 100 105
Gln Asp Lys Asn Leu Met Ala Gln Leu Gln Glu Thr Met Arg Cys
110 115 120
Val Cys Arg Phe Asp Asn Arg Thr Cys Arg Lys Leu Leu Ala Ser
125 130 135
Ile Ala Glu Asp Tyr Arg Lys Arg Ala Pro Tyr Ile Ala Tyr Leu
140 145 150
Thr Arg Cys Arg Gln Gly Leu Gln Thr Thr Gln Ala His Leu Glu
155 160 165
Arg Leu Leu Gln Arg Val Leu Arg Asp Lys Glu Val Ala Asn Arg
170 175 180
Tyr Phe Thr Thr Val Cys Val Arg Leu Leu Leu Glu Ser Lys Glu
185 190 195
Lys Lys Ile Arg Glu Phe Ile Gln Asp Phe Gln Lys Leu Thr Ala
200 205 210
Ala Asp Asp Lys Thr Ala Gln Val Glu Asp Phe Leu Gln Phe Leu

215 220 225
Tyr Gly Ala Met Ala Gln Asp Val Ile Trp Gln Asn Ala Ser Glu
230 235 240
Glu Gln Leu Gln Asp Ala Gln Leu Ala Ile Glu Arg Ser Val Met
245 250 255
Asn Arg Ile Phe Lys Leu Ala Phe Tyr Pro Asn Gln Asp Gly Asp
260 265 270
Ile Leu Arg Asp Gln Val Leu His Glu His Ile Gln Arg Leu Ser
275 280 285
Lys Val Val Thr Ala Asn His Arg Ala Leu Gln Ile Pro Glu Val
290 295 300
Tyr Leu Arg Glu Ala Pro Trp Pro Ser Ala Gln Ser Glu Ile Arg
305 310 315
Thr Ile Ser Ala Tyr Lys Thr Pro Arg Asp Lys Val Gln Cys Ile
320 325 330
Leu Arg Met Cys Ser Thr Ile Met Asn Leu Leu Ser Leu Ala Asn
335 340 345
Glu Asp Ser Val Pro Gly Ala Asp Asp Phe Val Pro Val Leu Val
350 355 360
Phe Val Leu Ile Lys Ala Asn Pro Pro Cys Leu Leu Ser Thr Val
365 370 375
Gln Tyr Ile Ser Ser Phe Tyr Ala Ser Cys Leu Ser Gly Glu Glu
380 385 390
Ser Tyr Trp Trp Met Gln Phe Thr Ala Ala Val Glu Phe Ile Lys
395 400 405
Thr Ile Asp Asp Arg Lys
410

<210> 11
<211> 201
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2374186CD1

<400> 11
Met Phe Pro Glu Gln Gln Lys Glu Glu Phe Val Ser Val Trp Val
1 5 10 15
Arg Asp Pro Arg Ile Gln Lys Glu Asp Phe Trp His Ser Tyr Ile
20 25 30
Asp Tyr Glu Ile Cys Ile His Thr Asn Ser Met Cys Phe Thr Met
35 40 45
Lys Thr Ser Cys Val Arg Arg Arg Tyr Arg Glu Phe Val Trp Leu
50 55 60
Arg Gln Arg Leu Gln Ser Asn Ala Leu Leu Val Gln Leu Pro Glu
65 70 75
Leu Pro Ser Lys Asn Leu Phe Phe Asn Met Asn Asn Arg Gln His
80 85 90
Val Asp Gln Arg Arg Gln Gly Leu Glu Asp Phe Leu Arg Lys Val
95 100 105
Leu Gln Asn Ala Leu Leu Leu Ser Asp Ser Ser Leu His Leu Phe
110 115 120
Leu Gln Ser His Leu Asn Ser Glu Asp Ile Glu Ala Cys Val Ser
125 130 135
Gly Gln Thr Lys Tyr Ser Val Glu Glu Ala Ile His Lys Phe Ala
140 145 150
Leu Met Asn Arg Arg Phe Pro Glu Glu Asp Glu Glu Gly Lys Lys
155 160 165
Glu Asn Asp Ile Asp Tyr Asp Ser Glu Ser Ser Ser Ser Gly Leu
170 175 180
Gly His Ser Ser Asp Asp Ser Ser Ser His Gly Cys Lys Val Asn
185 190 195
Thr Ala Pro Gln Glu Ser
200

<210> 12
<211> 476
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2476232CD1

<400> 12
Met Ala Phe Gly Lys Ser His Arg Asp Pro Tyr Ala Thr Ser Val
1 5 10 15
Gly His Leu Ile Glu Lys Ala Thr Phe Ala Gly Val Gln Thr Glu
20 25 30
Asp Trp Gly Gln Phe Met His Ile Cys Asp Ile Ile Asn Thr Thr
35 40 45
Gln Asp Gly Pro Lys Asp Ala Val Lys Ala Leu Lys Lys Arg Ile
50 55 60
Ser Lys Asn Tyr Asn His Lys Glu Ile Gln Leu Thr Leu Ser Leu
65 70 75
Ile Asp Met Cys Val Gln Asn Cys Gly Pro Ser Phe Gln Ser Leu
80 85 90
Ile Val Lys Lys Glu Phe Val Lys Glu Asn Leu Val Lys Leu Leu
95 100 105
Asn Pro Arg Tyr Asn Leu Pro Leu Asp Ile Gln Asn Arg Ile Leu
110 115 120
Asn Phe Ile Lys Thr Trp Ser Gln Gly Phe Pro Gly Gly Val Asp
125 130 135
Val Ser Glu Val Lys Glu Val Tyr Leu Asp Leu Val Lys Lys Gly
140 145 150
Val Gln Phe Pro Pro Ser Glu Ala Glu Ala Glu Thr Ala Arg Gln
155 160 165
Glu Thr Ala Gln Ile Ser Ser Asn Pro Pro Thr Ser Val Pro Thr
170 175 180
Ala Pro Ala Leu Ser Ser Val Ile Ala Pro Lys Asn Ser Thr Val
185 190 195
Thr Leu Val Pro Glu Gln Ile Gly Lys Leu His Ser Glu Leu Asp
200 205 210
Met Val Lys Met Asn Val Arg Val Met Ser Ala Ile Leu Met Glu
215 220 225
Asn Thr Pro Gly Ser Glu Asn His Glu Asp Ile Glu Leu Leu Gln
230 235 240
Lys Leu Tyr Lys Thr Gly Arg Glu Met Gln Glu Arg Ile Met Asp
245 250 255
Leu Leu Val Val Val Glu Asn Glu Asp Val Thr Val Glu Leu Ile
260 265 270
Gln Val Asn Glu Asp Leu Asn Asn Ala Ile Leu Gly Tyr Glu Arg
275 280 285
Phe Thr Arg Asn Gln Gln Arg Ile Leu Glu Gln Asn Lys Asn Gln
290 295 300
Lys Glu Ala Thr Asn Thr Thr Ser Glu Pro Ser Ala Pro Ser Gln
305 310 315
Asp Leu Leu Asp Leu Ser Pro Ser Pro Arg Met Pro Arg Ala Thr
320 325 330
Leu Gly Glu Leu Asn Thr Met Asn Asn Gln Leu Ser Gly Leu Asn
335 340 345
Phe Ser Leu Pro Ser Ser Asp Val Thr Asn Asn Leu Lys Pro Ser
350 355 360
Leu His Pro Gln Met Asn Leu Leu Ala Leu Glu Asn Thr Glu Ile
365 370 375
Pro Pro Phe Ala Gln Arg Thr Ser Gln Asn Leu Thr Ser Ser His
380 385 390
Ala Tyr Asp Asn Phe Leu Glu His Ser Asn Ser Val Phe Leu Gln
395 400 405
Pro Val Ser Leu Gln Thr Ile Ala Ala Ala Pro Ser Asn Gln Ser
410 415 420
Leu Pro Pro Leu Pro Ser Asn His Pro Ala Met Thr Lys Ser Asp

```
                425                 430                 435
Leu Gln Pro Pro Asn Tyr Tyr Glu Val Met Glu Phe Asp Pro Leu
                440                 445                 450
Ala Pro Ala Val Thr Thr Glu Ala Ile Tyr Glu Glu Ile Asp Ala
                455                 460                 465
His Gln His Lys Gly Ala Gln Asn Asp Gly Asp
                470                 475
```

<210> 13
<211> 1220
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2503986CD1

<400> 13

```
Met Lys Leu Lys Glu Val Asp Arg Thr Ala Met Gln Ala Trp Ser
  1               5                  10                  15
Pro Ala Gln Asn His Pro Ile Tyr Leu Ala Thr Gly Thr Ser Ala
             20                  25                  30
Gln Gln Leu Asp Ala Thr Phe Ser Thr Asn Ala Ser Leu Glu Ile
             35                  40                  45
Phe Glu Leu Asp Leu Ser Asp Pro Ser Leu Asp Met Lys Ser Cys
             50                  55                  60
Ala Thr Phe Ser Ser Ser His Arg Tyr His Lys Leu Ile Trp Gly
             65                  70                  75
Pro Tyr Lys Met Asp Ser Lys Gly Asp Val Ser Gly Val Leu Ile
             80                  85                  90
Ala Gly Gly Glu Asn Gly Asn Ile Ile Leu Tyr Asp Pro Ser Lys
             95                 100                 105
Ile Ile Ala Gly Asp Lys Glu Val Val Ile Ala Gln Asn Asp Lys
            110                 115                 120
His Thr Gly Pro Val Arg Ala Leu Asp Val Asn Ile Phe Gln Thr
            125                 130                 135
Asn Leu Val Ala Ser Gly Ala Asn Glu Ser Glu Ile Tyr Ile Trp
            140                 145                 150
Asp Leu Asn Asn Phe Ala Thr Pro Met Thr Pro Gly Ala Lys Thr
            155                 160                 165
Gln Pro Pro Glu Asp Ile Ser Cys Ile Ala Trp Asn Arg Gln Val
            170                 175                 180
Gln His Ile Leu Ala Ser Ala Ser Pro Ser Gly Arg Ala Thr Val
            185                 190                 195
Trp Asp Leu Arg Lys Asn Glu Pro Ile Ile Lys Val Ser Asp His
            200                 205                 210
Ser Asn Arg Met His Cys Ser Gly Leu Ala Trp His Pro Asp Val
            215                 220                 225
Ala Thr Gln Met Val Leu Ala Ser Glu Asp Asp Arg Leu Pro Val
            230                 235                 240
Ile Gln Met Trp Asp Leu Arg Phe Ala Ser Ser Pro Leu Arg Val
            245                 250                 255
Leu Glu Asn His Ala Arg Gly Ile Leu Ala Ile Ala Trp Ser Met
            260                 265                 270
Ala Asp Pro Glu Leu Leu Leu Ser Cys Gly Lys Asp Ala Lys Ile
            275                 280                 285
Leu Cys Ser Asn Pro Asn Thr Gly Glu Val Leu Tyr Glu Leu Pro
            290                 295                 300
Thr Asn Thr Gln Trp Cys Phe Asp Ile Gln Trp Cys Pro Arg Asn
            305                 310                 315
Pro Ala Val Leu Ser Ala Ala Ser Phe Asp Gly Arg Ile Ser Val
            320                 325                 330
Tyr Ser Ile Met Gly Gly Ser Thr Asp Gly Leu Arg Gln Lys Gln
            335                 340                 345
Val Asp Lys Leu Ser Ser Ser Phe Gly Asn Leu Asp Pro Phe Gly
            350                 355                 360
Thr Gly Gln Pro Leu Pro Pro Leu Gln Ile Pro Gln Gln Thr Ala
```

365 370 375
Gln His Ser Ile Val Leu Pro Leu Lys Lys Pro Pro Lys Trp Ile
380 385 390
Arg Arg Pro Val Gly Ala Ser Phe Ser Phe Gly Gly Lys Leu Val
395 400 405
Thr Phe Glu Asn Val Arg Met Pro Ser His Gln Gly Ala Glu Gln
410 415 420
Gln Gln Gln Gln His His Val Phe Ile Ser Gln Val Val Thr Glu
425 430 435
Lys Glu Phe Leu Ser Arg Ser Asp Gln Leu Gln Gln Ala Val Gln
440 445 450
Ser Gln Gly Phe Ile Asn Tyr Cys Gln Lys Lys Ile Asp Ala Ser
455 460 465
Gln Thr Glu Phe Glu Lys Asn Val Trp Ser Phe Leu Lys Val Asn
470 475 480
Phe Glu Asp Asp Ser Arg Gly Lys Tyr Leu Glu Leu Leu Gly Tyr
485 490 495
Arg Lys Glu Asp Leu Gly Lys Lys Ile Ala Leu Ala Leu Asn Lys
500 505 510
Val Asp Gly Ala Asn Val Ala Leu Lys Asp Ser Asp Gln Val Ala
515 520 525
Gln Ser Asp Gly Glu Glu Ser Pro Ala Ala Glu Glu Gln Leu Leu
530 535 540
Gly Glu His Ile Lys Glu Glu Lys Glu Glu Ser Glu Phe Leu Pro
545 550 555
Ser Ser Gly Gly Thr Phe Asn Ile Ser Val Ser Gly Asp Ile Asp
560 565 570
Gly Leu Ile Thr Gln Ala Leu Leu Thr Gly Asn Phe Glu Ser Ala
575 580 585
Val Asp Leu Cys Leu His Asp Asn Arg Met Ala Asp Ala Ile Ile
590 595 600
Leu Ala Ile Ala Gly Gly Gln Glu Leu Leu Ala Arg Thr Gln Lys
605 610 615
Lys Tyr Phe Ala Lys Ser Gln Ser Lys Ile Thr Arg Leu Ile Thr
620 625 630
Ala Val Val Met Lys Asn Trp Lys Glu Ile Val Glu Ser Cys Asp
635 640 645
Leu Lys Asn Trp Arg Glu Ala Leu Ala Ala Val Leu Thr Tyr Ala
650 655 660
Lys Pro Asp Glu Phe Ser Ala Leu Cys Asp Leu Leu Gly Thr Arg
665 670 675
Leu Glu Asn Glu Gly Asp Ser Leu Leu Gln Thr Gln Ala Cys Leu
680 685 690
Cys Tyr Ile Cys Ala Gly Asn Val Glu Lys Leu Val Ala Cys Trp
695 700 705
Thr Lys Ala Gln Asp Gly Ser His Pro Leu Ser Leu Gln Asp Leu
710 715 720
Ile Glu Lys Val Val Ile Leu Arg Lys Ala Val Gln Leu Thr Gln
725 730 735
Ala Met Asp Thr Ser Thr Val Gly Val Leu Leu Ala Ala Lys Met
740 745 750
Ser Gln Tyr Ala Asn Leu Leu Ala Ala Gln Gly Ser Ile Ala Ala
755 760 765
Ala Leu Ala Phe Leu Pro Asp Asn Thr Asn Gln Pro Asn Ile Met
770 775 780
Gln Leu Arg Asp Arg Leu Cys Arg Ala Gln Gly Glu Pro Val Ala
785 790 795
Gly His Glu Ser Pro Lys Ile Pro Tyr Glu Lys Gln Gln Leu Pro
800 805 810
Lys Gly Arg Pro Gly Pro Val Ala Gly His His Gln Met Pro Arg
815 820 825
Val Gln Thr Gln Gln Tyr Tyr Pro His Gly Glu Asn Pro Pro Pro
830 835 840
Pro Gly Phe Ile Met His Gly Asn Val Asn Pro Asn Ala Ala Gly
845 850 855
Gln Leu Pro Thr Ser Pro Gly His Met His Thr Gln Val Pro Pro
860 865 870

```
Tyr Pro Gln Pro Gln Pro Tyr Gln Pro Ala Gln Pro Tyr Pro Phe
                875                 880                 885
Gly Thr Gly Gly Ser Ala Met Tyr Arg Pro Gln Gln Pro Val Ala
                890                 895                 900
Pro Pro Thr Ser Asn Ala Tyr Pro Asn Thr Pro Tyr Ile Ser Ser
                905                 910                 915
Ala Ser Ser Tyr Thr Gly Gln Ser Gln Leu Tyr Ala Ala Gln His
                920                 925                 930
Gln Ala Ser Ser Pro Thr Ser Ser Pro Ala Thr Ser Phe Pro Pro
                935                 940                 945
Pro Pro Ser Ser Gly Ala Ser Phe Gln His Gly Gly Pro Gly Ala
                950                 955                 960
Pro Pro Ser Ser Ser Ala Tyr Ala Leu Pro Pro Gly Thr Thr Gly
                965                 970                 975
Thr Leu Pro Ala Ala Ser Glu Leu Pro Ala Ser Gln Arg Thr Gly
                980                 985                 990
Pro Gln Asn Gly Trp Asn Asp Pro Pro Ala Leu Asn Arg Val Pro
                995                1000                1005
Lys Lys Lys Lys Met Pro Glu Asn Phe Met Pro Pro Val Pro Ile
               1010                1015                1020
Thr Ser Pro Ile Met Asn Pro Leu Gly Asp Pro Gln Ser Gln Met
               1025                1030                1035
Leu Gln Gln Gln Pro Ser Ala Pro Val Pro Leu Ser Ser Gln Ser
               1040                1045                1050
Ser Phe Pro Gln Pro His Leu Pro Gly Gly Gln Pro Phe His Gly
               1055                1060                1065
Val Gln Gln Pro Leu Gly Gln Thr Gly Met Pro Pro Ser Phe Ser
               1070                1075                1080
Lys Pro Asn Ile Glu Gly Ala Pro Gly Ala Pro Ile Gly Asn Thr
               1085                1090                1095
Phe Gln His Val Gln Ser Leu Pro Thr Lys Lys Ile Thr Lys Lys
               1100                1105                1110
Pro Ile Pro Asp Glu His Leu Ile Leu Lys Thr Thr Phe Glu Asp
               1115                1120                1125
Leu Ile Gln Arg Cys Leu Ser Ser Ala Thr Asp Pro Gln Thr Lys
               1130                1135                1140
Arg Lys Leu Asp Asp Ala Ser Lys Arg Leu Glu Phe Leu Tyr Asp
               1145                1150                1155
Lys Leu Arg Glu Gln Thr Leu Ser Pro Thr Ile Thr Ser Gly Leu
               1160                1165                1170
His Asn Ile Ala Arg Ser Ile Glu Thr Arg Asn Tyr Ser Glu Gly
               1175                1180                1185
Leu Thr Met His Thr His Ile Val Ser Thr Ser Asn Phe Ser Glu
               1190                1195                1200
Thr Ser Ala Phe Met Pro Val Leu Lys Val Val Leu Thr Gln Ala
               1205                1210                1215
Asn Lys Leu Gly Val
               1220

<210> 14
<211> 222
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2596566CD1

<400> 14
Met Ala Phe Val Lys Ser Gly Trp Leu Leu Arg Gln Ser Thr Ile
  1               5                  10                  15
Leu Lys Arg Trp Lys Lys Asn Trp Phe Asp Leu Trp Ser Asp Gly
             20                  25                  30
His Leu Ile Tyr Tyr Asp Asp Gln Thr Arg Gln Asn Ile Glu Asp
             35                  40                  45
Lys Val His Met Pro Met Asp Cys Ile Asn Ile Arg Thr Gly Gln
             50                  55                  60
```

Glu Cys Arg Asp Thr Gln Pro Pro Asp Gly Lys Ser Lys Asp Cys
65 70 75
Met Leu Gln Ile Val Cys Arg Asp Gly Lys Thr Ile Ser Leu Cys
80 85 90
Ala Glu Ser Thr Asp Asp Cys Leu Ala Trp Lys Phe Thr Leu Gln
95 100 105
Asp Ser Arg Thr Asn Thr Ala Tyr Val Gly Ser Ala Val Met Thr
110 115 120
Asp Glu Thr Ser Val Val Ser Ser Pro Pro Pro Tyr Thr Ala Tyr
125 130 135
Ala Ala Pro Ala Pro Glu Gln Ala Tyr Gly Tyr Gly Pro Tyr Gly
140 145 150
Gly Ala Tyr Pro Pro Gly Thr Gln Val Val Tyr Ala Ala Asn Gly
155 160 165
Gln Ala Tyr Ala Val Pro Tyr Gln Tyr Pro Tyr Ala Gly Leu Tyr
170 175 180
Gly Gln Gln Pro Ala Asn Gln Val Ile Ile Arg Glu Arg Tyr Arg
185 190 195
Asp Asn Asp Ser Asp Leu Ala Leu Gly Met Leu Ala Gly Ala Ala
200 205 210
Thr Gly Met Ala Leu Gly Ser Leu Phe Trp Val Phe
215 220

<210> 15
<211> 924
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2685253CD1

<400> 15
Met Ser Arg Ser Arg Gln Pro Pro Leu Val Thr Gly Ile Ser Pro
1 5 10 15
Asn Glu Gly Ile Pro Trp Thr Lys Val Thr Ile Arg Gly Glu Asn
20 25 30
Leu Gly Thr Gly Pro Thr Asp Leu Ile Gly Leu Thr Ile Cys Gly
35 40 45
His Asn Cys Leu Leu Thr Ala Glu Trp Met Ser Ala Ser Lys Ile
50 55 60
Val Cys Arg Val Gly Gln Ala Lys Asn Asp Lys Gly Asp Ile Ile
65 70 75
Val Thr Thr Lys Ser Gly Gly Arg Gly Thr Ser Thr Val Ser Phe
80 85 90
Lys Leu Leu Lys Pro Glu Lys Ile Gly Ile Leu Asp Gln Ser Ala
95 100 105
Val Trp Val Asp Glu Met Asn Tyr Tyr Asp Met Arg Thr Asp Arg
110 115 120
Asn Lys Gly Ile Pro Pro Leu Ser Leu Arg Pro Ala Asn Pro Leu
125 130 135
Gly Ile Glu Ile Glu Lys Ser Lys Phe Ser Gln Lys Asp Leu Glu
140 145 150
Met Leu Phe His Gly Met Ser Ala Asp Phe Thr Ser Glu Asn Phe
155 160 165
Ser Ala Ala Trp Tyr Leu Ile Glu Asn His Ser Asn Thr Ser Phe
170 175 180
Glu Gln Leu Lys Met Ala Val Thr Asn Leu Lys Arg Gln Ala Asn
185 190 195
Lys Lys Ser Glu Gly Ser Leu Ala Tyr Val Lys Gly Gly Leu Ser
200 205 210
Thr Phe Phe Glu Ala Gln Asp Ala Leu Ser Ala Ile His Gln Lys
215 220 225
Leu Glu Ala Asp Gly Thr Glu Lys Val Glu Gly Ser Met Thr Gln
230 235 240
Lys Leu Glu Asn Val Leu Asn Arg Ala Ser Asn Thr Ala Asp Thr
245 250 255

Leu Phe Gln Glu Val Leu Gly Arg Lys Asp Lys Ala Asp Ser Thr
260 265 270
Arg Asn Ala Leu Asn Val Leu Gln Arg Phe Lys Phe Leu Phe Asn
275 280 285
Leu Pro Leu Asn Ile Glu Arg Asn Ile Gln Lys Gly Asp Tyr Asp
290 295 300
Val Val Ile Asn Asp Tyr Glu Lys Ala Lys Ser Leu Phe Gly Lys
305 310 315
Thr Glu Val Gln Val Phe Lys Lys Tyr Tyr Ala Glu Val Glu Thr
320 325 330
Arg Ile Glu Ala Leu Arg Glu Leu Leu Leu Asp Lys Leu Leu Glu
335 340 345
Thr Pro Ser Thr Leu His Asp Gln Lys Arg Tyr Ile Arg Tyr Leu
350 355 360
Ser Asp Leu His Ala Ser Gly Asp Pro Ala Trp Gln Cys Ile Gly
365 370 375
Ala Gln His Lys Trp Ile Leu Gln Leu Met His Ser Cys Lys Glu
380 385 390
Gly Tyr Val Lys Asp Leu Lys Gly Asn Pro Gly Leu His Ser Pro
395 400 405
Met Leu Asp Leu Asp Asn Asp Thr Arg Pro Ser Val Leu Gly His
410 415 420
Leu Ser Gln Thr Ala Ser Leu Lys Arg Gly Ser Ser Phe Gln Ser
425 430 435
Gly Arg Asp Asp Thr Trp Arg Tyr Lys Thr Pro His Arg Val Ala
440 445 450
Phe Val Glu Lys Leu Thr Lys Leu Val Leu Ser Gln Leu Pro Asn
455 460 465
Phe Trp Lys Leu Trp Ile Ser Tyr Val Asn Gly Ser Leu Phe Ser
470 475 480
Glu Thr Ala Glu Lys Ser Gly Gln Ile Glu Arg Ser Lys Asn Val
485 490 495
Arg Gln Arg Gln Asn Asp Phe Lys Lys Met Ile Gln Glu Val Met
500 505 510
His Ser Leu Val Lys Leu Thr Arg Gly Ala Leu Leu Pro Leu Ser
515 520 525
Ile Arg Asp Gly Glu Ala Lys Gln Tyr Gly Gly Trp Glu Val Lys
530 535 540
Cys Glu Leu Ser Gly Gln Trp Leu Ala His Ala Ile Gln Thr Val
545 550 555
Arg Leu Thr His Glu Ser Leu Thr Ala Leu Glu Ile Pro Asn Asp
560 565 570
Leu Leu Gln Thr Ile Gln Asp Leu Ile Leu Asp Leu Arg Val Arg
575 580 585
Cys Val Met Ala Thr Leu Gln His Thr Ala Glu Glu Ile Lys Arg
590 595 600
Leu Ala Glu Lys Glu Asp Trp Ile Val Asp Asn Glu Gly Leu Thr
605 610 615
Ser Leu Pro Cys Gln Phe Glu Gln Cys Ile Val Cys Ser Leu Gln
620 625 630
Ser Leu Lys Gly Val Leu Glu Cys Lys Pro Gly Glu Ala Ser Val
635 640 645
Phe Gln Gln Pro Lys Thr Gln Glu Glu Val Cys Gln Leu Ser Ile
650 655 660
Asn Ile Met Gln Val Phe Ile Tyr Cys Leu Glu Gln Leu Ser Thr
665 670 675
Lys Pro Asp Ala Asp Ile Asp Thr Thr His Leu Ser Val Asp Val
680 685 690
Ser Ser Pro Asp Leu Phe Gly Ser Ile His Glu Asp Phe Ser Leu
695 700 705
Thr Ser Glu Gln Arg Leu Leu Ile Val Leu Ser Asn Cys Cys Tyr
710 715 720
Leu Glu Arg His Thr Phe Leu Asn Ile Ala Glu His Phe Glu Lys
725 730 735
His Asn Phe Gln Gly Ile Glu Lys Ile Thr Gln Val Ser Met Ala
740 745 750
Ser Leu Lys Glu Leu Asp Gln Arg Leu Phe Glu Asn Tyr Ile Glu

```
                755                 760                 765
Leu Lys Ala Asp Pro Ile Val Gly Ser Leu Glu Pro Gly Ile Tyr
            770                 775                 780
Ala Gly Tyr Phe Asp Trp Lys Asp Cys Leu Pro Pro Thr Gly Val
        785                 790                 795
Arg Asn Tyr Leu Lys Glu Ala Leu Val Asn Ile Ile Ala Val His
    800                 805                 810
Ala Glu Val Phe Thr Ile Ser Lys Glu Leu Val Pro Arg Val Leu
815                 820                 825
Ser Lys Val Ile Glu Ala Val Ser Glu Glu Leu Ser Arg Leu Met
            830                 835                 840
Gln Cys Val Ser Ser Phe Ser Lys Asn Gly Ala Leu Gln Ala Arg
        845                 850                 855
Leu Glu Ile Cys Ala Leu Arg Asp Thr Val Ala Val Tyr Leu Thr
    860                 865                 870
Pro Glu Ser Lys Ser Ser Phe Lys Gln Ala Leu Glu Ala Leu Pro
875                 880                 885
Gln Leu Ser Ser Gly Ala Asp Lys Lys Leu Leu Glu Glu Leu Leu
            890                 895                 900
Asn Lys Phe Lys Ser Ser Met His Leu Gln Leu Thr Cys Phe Gln
        905                 910                 915
Ala Ala Ser Ser Thr Met Met Lys Thr
    920
```

<210> 16
<211> 435
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2762252CD1

<400> 16

```
Met Ala Pro Phe Gly Arg Asn Leu Leu Lys Thr Arg His Lys Asn
 1               5                  10                  15
Arg Ser Pro Thr Lys Asp Met Asp Ser Glu Glu Lys Glu Ile Val
            20                  25                  30
Val Trp Val Cys Gln Glu Glu Lys Leu Val Cys Gly Leu Thr Lys
        35                  40                  45
Arg Thr Thr Ser Ala Asp Val Ile Gln Ala Leu Leu Glu Glu His
    50                  55                  60
Glu Ala Thr Phe Gly Glu Lys Arg Phe Leu Leu Gly Lys Pro Ser
65                  70                  75
Asp Tyr Cys Ile Ile Glu Lys Trp Arg Gly Ser Glu Arg Val Leu
            80                  85                  90
Pro Pro Leu Thr Arg Ile Leu Lys Leu Trp Lys Ala Trp Gly Asp
        95                  100                 105
Glu Gln Pro Asn Met Gln Phe Val Leu Val Lys Ala Asp Ala Phe
    110                 115                 120
Leu Pro Val Pro Leu Trp Arg Thr Ala Glu Ala Lys Leu Val Gln
125                 130                 135
Asn Thr Glu Lys Leu Trp Glu Leu Ser Pro Ala Asn Tyr Met Lys
            140                 145                 150
Thr Leu Pro Pro Asp Lys Gln Lys Arg Ile Val Arg Lys Thr Phe
        155                 160                 165
Arg Lys Leu Ala Lys Ile Lys Gln Asp Thr Val Ser His Asp Arg
    170                 175                 180
Asp Asn Met Glu Thr Leu Val His Leu Ile Ile Ser Gln Asp His
185                 190                 195
Thr Ile His Gln Gln Val Lys Arg Met Lys Glu Leu Asp Leu Glu
            200                 205                 210
Ile Glu Lys Cys Glu Ala Lys Phe His Leu Asp Arg Val Glu Asn
        215                 220                 225
Asp Gly Glu Asn Tyr Val Gln Asp Ala Tyr Leu Met Pro Ser Phe
    230                 235                 240
Ser Glu Val Glu Gln Asn Leu Asp Leu Gln Tyr Glu Glu Asn Gln
```

```
                245                 250                 255
Thr Leu Glu Asp Leu Ser Glu Ser Asp Gly Ile Glu Gln Leu Glu
            260                 265                 270
Glu Arg Leu Lys Tyr Tyr Arg Ile Leu Ile Asp Lys Leu Ser Ala
        275                 280                 285
Glu Ile Glu Lys Glu Val Lys Ser Val Cys Ile Asp Ile Asn Glu
    290                 295                 300
Asp Ala Glu Gly Glu Ala Ala Ser Glu Leu Glu Ser Ser Asn Leu
305                 310                 315
Glu Ser Val Lys Cys Asp Leu Glu Lys Ser Met Lys Ala Gly Leu
            320                 325                 330
Lys Ile His Ser His Leu Ser Gly Ile Gln Lys Glu Ile Lys Tyr
        335                 340                 345
Ser Asp Ser Leu Leu Gln Met Lys Ala Lys Glu Tyr Glu Leu Leu
    350                 355                 360
Ala Lys Glu Phe Asn Ser Leu His Ile Ser Asn Lys Asp Gly Cys
365                 370                 375
Gln Leu Lys Glu Asn Arg Ala Lys Glu Ser Glu Val Pro Ser Ser
            380                 385                 390
Asn Gly Glu Ile Pro Pro Phe Thr Gln Arg Val Phe Ser Asn Tyr
        395                 400                 405
Thr Asn Asp Thr Asp Ser Asp Thr Gly Ile Ser Ser Asn His Ser
    410                 415                 420
Gln Asp Ser Glu Thr Thr Val Gly Asp Val Val Leu Leu Ser Thr
425                 430                 435
```

<210> 17
<211> 321
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3452009CD1

<400> 17

```
Met Glu Phe Leu Leu Gly Asn Pro Phe Ser Thr Pro Val Gly Gln
  1               5                  10                  15
Cys Leu Glu Lys Ala Thr Asp Gly Ser Leu Gln Ser Glu Asp Trp
            20                  25                  30
Thr Leu Asn Met Glu Ile Cys Asp Ile Ile Asn Glu Thr Glu Glu
        35                  40                  45
Gly Pro Lys Asp Ala Ile Arg Ala Leu Lys Lys Arg Leu Asn Gly
    50                  55                  60
Asn Arg Asn Tyr Arg Glu Val Met Leu Ala Leu Thr Val Leu Glu
65                  70                  75
Thr Cys Val Lys Asn Cys Gly His Arg Phe His Ile Leu Val Ala
            80                  85                  90
Asn Arg Asp Phe Ile Asp Ser Val Leu Val Lys Ile Ile Ser Pro
        95                  100                 105
Lys Asn Asn Pro Pro Thr Ile Val Gln Asp Lys Val Leu Ala Leu
    110                 115                 120
Ile Gln Ala Trp Ala Asp Ala Phe Arg Ser Ser Pro Asp Leu Thr
125                 130                 135
Gly Val Val His Ile Tyr Glu Glu Leu Lys Arg Lys Gly Val Glu
            140                 145                 150
Phe Pro Met Ala Asp Leu Asp Ala Leu Ser Pro Ile His Thr Pro
        155                 160                 165
Gln Arg Ser Val Pro Glu Val Asp Pro Ala Ala Thr Met Pro Arg
    170                 175                 180
Ser Gln Ser Gln Gln Arg Thr Ser Ala Gly Ser Tyr Ser Ser Pro
185                 190                 195
Pro Pro Ala Pro Tyr Ser Ala Pro Gln Ala Pro Ala Leu Ser Val
            200                 205                 210
Thr Gly Pro Ile Thr Ala Asn Ser Glu Gln Ile Ala Arg Leu Arg
        215                 220                 225
```

Ser Glu Leu Asp Val Val Arg Gly Asn Thr Lys Val Met Ser Glu
230 235 240
Met Leu Thr Glu Met Val Pro Gly Gln Glu Asp Ser Ser Asp Leu
245 250 255
Glu Leu Leu Gln Glu Leu Asn Arg Thr Cys Arg Ala Met Gln Gln
260 265 270
Arg Ile Val Glu Leu Ile Ser Arg Val Ser Asn Glu Glu Val Thr
275 280 285
Glu Glu Leu Leu His Val Asn Asp Asp Leu Asn Asn Val Phe Leu
290 295 300
Arg Tyr Glu Arg Trp Glu Pro Asp Phe Phe Phe Phe Phe Phe Pro
305 310 315
Leu Lys Arg Leu Leu Pro
320

<210> 18
<211> 499
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4644780CD1

<400> 18
Met Pro Ala Val Ser Gly Pro Gly Pro Leu Phe Cys Leu Leu Leu
1 5 10 15
Leu Leu Leu Asp Pro His Ser Pro Glu Thr Gly Cys Pro Pro Leu
20 25 30
Arg Arg Phe Glu Tyr Lys Leu Ser Phe Lys Gly Pro Arg Leu Ala
35 40 45
Leu Pro Gly Ala Gly Ile Pro Phe Trp Ser His His Gly Asp Ala
50 55 60
Ile Leu Gly Leu Glu Glu Val Arg Leu Thr Pro Ser Met Arg Asn
65 70 75
Arg Ser Gly Ala Val Trp Ser Arg Ala Ser Val Pro Phe Ser Ala
80 85 90
Trp Glu Val Glu Val Gln Met Arg Val Thr Gly Leu Gly Arg Arg
95 100 105
Gly Ala Gln Gly Met Ala Val Trp Tyr Thr Arg Gly Arg Gly His
110 115 120
Val Gly Ser Val Leu Gly Gly Leu Ala Ser Trp Asp Gly Ile Gly
125 130 135
Ile Phe Phe Asp Ser Pro Ala Glu Asp Thr Gln Asp Ser Pro Ala
140 145 150
Ile Arg Val Leu Ala Ser Asp Gly His Ile Pro Ser Glu Gln Pro
155 160 165
Gly Asp Gly Ala Ser Gln Gly Leu Gly Ser Cys His Trp Asp Phe
170 175 180
Arg Asn Arg Pro His Pro Phe Arg Ala Arg Ile Thr Tyr Trp Gly
185 190 195
Gln Arg Leu Arg Met Ser Leu Asn Ser Gly Leu Thr Pro Ser Asp
200 205 210
Pro Asp Asp His Asp Val Leu Ser Phe Leu Thr Phe Ser Leu Ser
215 220 225
Glu Pro Ser Pro Glu Val Pro Pro Gln Pro Phe Leu Glu Met Gln
230 235 240
Gln Leu Arg Leu Ala Arg Gln Leu Glu Gly Leu Trp Ala Arg Leu
245 250 255
Gly Leu Gly Thr Arg Glu Asp Val Thr Pro Lys Ser Asp Ser Glu
260 265 270
Ala Gln Gly Glu Gly Glu Arg Leu Phe Asp Leu Glu Glu Thr Leu
275 280 285
Gly Arg His Arg Arg Ile Leu Gln Ala Leu Arg Gly Leu Ser Lys
290 295 300
Gln Leu Ala Gln Ala Glu Arg Gln Trp Lys Lys Gln Leu Gly Pro
305 310 315

```
Pro Gly Gln Ala Arg Pro Asp Gly Gly Trp Ala Leu Asp Ala Ser
                320                 325                 330
Cys Gln Ile Pro Ser Thr Pro Gly Arg Gly Gly His Leu Ser Met
                335                 340                 345
Ser Leu Asn Lys Asp Ser Ala Lys Val Gly Ala Leu Leu His Gly
                350                 355                 360
Gln Trp Thr Leu Leu Gln Ala Leu Gln Glu Met Arg Asp Ala Ala
                365                 370                 375
Val Arg Met Ala Ala Glu Ala Gln Val Ser Tyr Leu Pro Val Gly
                380                 385                 390
Ile Glu His His Phe Leu Glu Leu Asp His Ile Leu Gly Leu Leu
                395                 400                 405
Gln Glu Glu Leu Arg Gly Pro Ala Lys Ala Ala Ala Lys Ala Pro
                410                 415                 420
Arg Pro Pro Gly Gln Pro Pro Arg Ala Ser Ser Cys Leu Gln Pro
                425                 430                 435
Gly Ile Phe Leu Phe Tyr Leu Leu Ile Gln Thr Val Gly Phe Phe
                440                 445                 450
Gly Tyr Val His Phe Ser Arg Gln Glu Leu Asn Lys Ser Leu Gln
                455                 460                 465
Glu Cys Leu Ser Thr Gly Ser Leu Pro Leu Gly Pro Ala Pro His
                470                 475                 480
Thr Pro Arg Ala Leu Gly Ile Leu Arg Arg Gln Pro Leu Pro Ala
                485                 490                 495
Ser Met Pro Ala
```

<210> 19
<211> 879
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4946103CD1

<400> 19

```
Met Thr Ala Ile Lys His Ala Leu Gln Arg Asp Ile Phe Thr Pro
  1                 5                  10                  15
Asn Asp Glu Arg Leu Leu Ser Ile Val Asn Val Cys Lys Ala Gly
                 20                  25                  30
Lys Lys Lys Lys Asn Cys Phe Leu Cys Ala Thr Val Thr Thr Glu
                 35                  40                  45
Arg Pro Val Gln Val Lys Val Val Lys Val Lys Lys Ser Asp Lys
                 50                  55                  60
Gly Asp Phe Tyr Lys Arg Gln Ile Ala Trp Ala Leu Arg Asp Leu
                 65                  70                  75
Ala Val Val Asp Ala Lys Asp Ala Ile Lys Glu Asn Pro Glu Phe
                 80                  85                  90
Asp Leu His Phe Glu Lys Ile Tyr Lys Trp Val Ala Ser Ser Thr
                 95                 100                 105
Ala Glu Lys Asn Ala Phe Ile Ser Cys Ile Trp Lys Leu Asn Gln
                110                 115                 120
Arg Tyr Leu Arg Lys Lys Ile Asp Phe Val Asn Val Ser Ser Gln
                125                 130                 135
Leu Leu Glu Glu Ser Val Pro Ser Gly Glu Asn Gln Ser Val Thr
                140                 145                 150
Gly Gly Asp Glu Glu Val Val Asp Glu Tyr Gln Glu Leu Asn Ala
                155                 160                 165
Arg Glu Glu Gln Asp Ile Glu Ile Met Met Glu Gly Cys Glu Tyr
                170                 175                 180
Ala Ile Ser Asn Ala Glu Arg Phe Ala Glu Lys Leu Ser Arg Glu
                185                 190                 195
Leu Gln Val Leu Asp Gly Ala Asn Ile Gln Ser Ile Met Ala Ser
                200                 205                 210
Glu Lys Gln Val Asn Ile Leu Met Lys Leu Leu Asp Glu Ala Leu
                215                 220                 225
```

Lys Glu Val Asp Gln Ile Glu Leu Lys Leu Ser Ser Tyr Glu Glu
230 235 240
Met Leu Gln Ser Val Lys Glu Gln Met Asp Gln Ile Ser Glu Ser
245 250 255
Asn His Leu Ile His Leu Ser Asn Thr Asn Asn Val Lys Leu Leu
260 265 270
Ser Glu Ile Glu Phe Leu Val Asn His Met Asp Leu Ala Lys Gly
275 280 285
His Ile Lys Ala Leu Gln Glu Gly Asp Leu Ala Ser Ser Arg Gly
290 295 300
Ile Glu Ala Cys Thr Asn Ala Ala Asp Ala Leu Leu Gln Cys Met
305 310 315
Asn Val Ala Leu Arg Pro Gly His Asp Leu Leu Leu Ala Val Lys
320 325 330
Gln Gln Gln Gln Arg Phe Ser Asp Leu Arg Glu Leu Phe Ala Arg
335 340 345
Arg Leu Ala Ser His Leu Asn Asn Val Phe Val Gln Gln Gly His
350 355 360
Asp Gln Ser Ser Ser Leu Pro Gln His Cys Val Ser Thr Gly Phe
365 370 375
Thr Gln Ser Ser Ser Ile Ser Gln Arg Phe Pro Pro Ile Ala Lys
380 385 390
Leu Met Glu Trp Leu Lys Ser Thr Asp Tyr Gly Lys Tyr Glu Gly
395 400 405
Leu Thr Lys Asn Tyr Met Asp Tyr Leu Ser Arg Leu Tyr Glu Arg
410 415 420
Glu Ile Lys Asp Phe Phe Glu Val Ala Lys Ile Lys Met Thr Gly
425 430 435
Thr Thr Lys Glu Ser Lys Lys Phe Gly Leu His Gly Ser Ser Gly
440 445 450
Lys Leu Thr Gly Ser Thr Ser Ser Leu Asn Lys Leu Ser Val Gln
455 460 465
Ser Ser Gly Asn Arg Arg Ser Gln Ser Ser Ser Leu Leu Asp Met
470 475 480
Gly Asn Met Ser Ala Ser Asp Leu Asp Val Ala Asp Arg Thr Lys
485 490 495
Phe Asp Lys Ile Phe Glu Gln Val Leu Ser Glu Leu Glu Pro Leu
500 505 510
Cys Leu Ala Glu Gln Asp Phe Ile Ser Lys Phe Phe Lys Leu Gln
515 520 525
Gln His Gln Ser Met Pro Gly Thr Met Ala Glu Ala Glu Asp Leu
530 535 540
Asp Gly Gly Thr Leu Ser Arg Gln His Asn Cys Gly Thr Pro Leu
545 550 555
Pro Val Ser Ser Glu Lys Asp Met Ile Arg Gln Met Met Ile Lys
560 565 570
Ile Phe Arg Cys Ile Glu Pro Glu Leu Asn Asn Leu Ile Ala Leu
575 580 585
Gly Asp Lys Ile Asp Ser Phe Asn Ser Leu Tyr Met Leu Val Lys
590 595 600
Met Ser His His Val Trp Thr Ala Gln Asn Val Asp Pro Ala Ser
605 610 615
Phe Leu Ser Thr Thr Leu Gly Asn Val Leu Val Thr Val Lys Arg
620 625 630
Asn Phe Asp Lys Cys Ile Ser Asn Gln Ile Arg Gln Met Glu Glu
635 640 645
Val Lys Ile Ser Lys Lys Ser Lys Val Gly Ile Leu Pro Phe Val
650 655 660
Ala Glu Phe Glu Glu Phe Ala Gly Leu Ala Glu Ser Ile Phe Lys
665 670 675
Asn Ala Glu Arg Arg Gly Asp Leu Asp Lys Ala Tyr Thr Lys Leu
680 685 690
Ile Arg Gly Val Phe Val Asn Val Glu Lys Val Ala Asn Glu Ser
695 700 705
Gln Lys Thr Pro Arg Asp Val Val Met Met Glu Asn Phe His His
710 715 720
Ile Phe Ala Thr Leu Ser Arg Leu Lys Ile Ser Cys Leu Glu Ala

725 730 735
Glu Lys Lys Glu Ala Lys Gln Lys Tyr Thr Asp His Leu Gln Ser
740 745 750
Tyr Val Ile Tyr Ser Leu Gly Gln Pro Leu Glu Lys Leu Asn His
755 760 765
Phe Phe Glu Gly Val Glu Ala Arg Val Ala Gln Gly Ile Arg Glu
770 775 780
Glu Glu Val Ser Tyr Gln Leu Ala Phe Asn Lys Gln Glu Leu Arg
785 790 795
Lys Val Ile Lys Glu Tyr Pro Gly Lys Glu Val Lys Lys Gly Leu
800 805 810
Asp Asn Leu Tyr Lys Lys Val Asp Lys His Leu Cys Glu Glu Glu
815 820 825
Asn Leu Leu Gln Val Val Trp His Ser Met Gln Asp Glu Phe Ile
830 835 840
Arg Gln Tyr Lys His Phe Glu Gly Leu Ile Ala Arg Cys Tyr Pro
845 850 855
Gly Ser Gly Val Thr Met Glu Phe Thr Ile Gln Asp Ile Leu Asp
860 865 870
Tyr Cys Ser Ser Ile Ala Gln Ser His
875

<210> 20
<211> 298
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5562355CD1

<400> 20
Met Asp Asn Ala Gly Lys Glu Arg Glu Ala Val Gln Leu Met Ala
1 5 10 15
Glu Ala Glu Lys Arg Val Lys Ala Ser His Ser Phe Leu Arg Gly
20 25 30
Leu Phe Gly Gly Asn Thr Arg Ile Glu Glu Ala Cys Glu Met Tyr
35 40 45
Thr Arg Ala Ala Asn Met Phe Lys Met Ala Lys Asn Trp Ser Ala
50 55 60
Ala Gly Asn Ala Phe Cys Gln Ala Ala Lys Leu His Met Gln Leu
65 70 75
Gln Ser Lys His Asp Ser Ala Thr Ser Phe Val Asp Ala Gly Asn
80 85 90
Ala Tyr Lys Lys Ala Asp Pro Gln Glu Ala Ile Asn Cys Leu Asn
95 100 105
Ala Ala Ile Asp Ile Tyr Thr Asp Met Gly Arg Phe Thr Ile Ala
110 115 120
Ala Lys His His Ile Thr Ile Ala Glu Ile Tyr Glu Thr Glu Leu
125 130 135
Val Asp Ile Glu Lys Ala Ile Ala His Tyr Glu Gln Ser Ala Asp
140 145 150
Tyr Tyr Lys Gly Glu Glu Ser Asn Ser Ser Ala Asn Lys Cys Leu
155 160 165
Leu Lys Val Ala Ala Tyr Ala Ala His Leu Glu Gln Tyr Gln Asn
170 175 180
Ala Ile Glu Ile Tyr Glu Gln Val Gly Ala Asn Thr Met Asp Asn
185 190 195
Pro Leu Thr Thr Tyr Ser Ala Lys Asp Tyr Phe Phe Lys Ala Ala
200 205 210
Leu Cys His Phe Ile Val Asp Glu Leu Asn Ala Lys Leu Ala Leu
215 220 225
Glu Gln Tyr Glu Asp Met Phe Pro Ala Phe Thr Asp Ser Arg Glu
230 235 240
Cys Lys Leu Leu Lys Lys Leu Leu Glu Ala His Glu Glu Gln Asn
245 250 255
Ser Glu Ala Tyr Thr Glu Ala Val Lys Glu Phe Asp Ser Ile Ser

260 265 270
Arg Leu Asp Gln Trp Leu Thr Thr Met Leu Leu Arg Ile Lys Lys
275 280 285
Ser Ile Gln Gly Asp Gly Glu Gly Asp Gly Asp Leu Lys
290 295

<210> 21
<211> 941
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5678824CD1

<400> 21
Met Ala Ala Tyr Leu Gln Trp Arg Arg Phe Val Phe Phe Asp Lys
1 5 10 15
Glu Leu Val Lys Glu Pro Leu Ser Asn Asp Gly Ala Ala Pro Gly
20 25 30
Ala Thr Pro Ala Ser Gly Ser Ala Ala Ser Lys Phe Leu Cys Leu
35 40 45
Pro Pro Gly Ile Thr Val Cys Asp Ser Gly Arg Gly Ser Leu Val
50 55 60
Phe Gly Asp Met Glu Gly Gln Ile Trp Phe Leu Pro Arg Ser Leu
65 70 75
Gln Leu Thr Gly Phe Gln Ala Tyr Lys Leu Arg Val Thr His Leu
80 85 90
Tyr Gln Leu Lys Gln His Asn Ile Leu Ala Ser Val Gly Glu Asp
95 100 105
Glu Glu Gly Ile Asn Pro Leu Val Lys Ile Trp Asn Leu Glu Lys
110 115 120
Arg Asp Gly Gly Asn Pro Leu Cys Thr Arg Ile Phe Pro Ala Ile
125 130 135
Pro Gly Thr Glu Pro Thr Val Val Ser Cys Leu Thr Val His Glu
140 145 150
Asn Leu Asn Phe Met Ala Ile Gly Phe Thr Asp Gly Ser Val Thr
155 160 165
Leu Asn Lys Gly Asp Ile Thr Arg Asp Arg His Ser Lys Thr Gln
170 175 180
Ile Leu His Lys Gly Asn Tyr Pro Val Thr Gly Leu Ala Phe Arg
185 190 195
Gln Ala Gly Lys Thr Thr His Leu Phe Val Val Thr Thr Glu Asn
200 205 210
Val Gln Ser Tyr Ile Val Ser Gly Lys Asp Tyr Pro Arg Val Glu
215 220 225
Leu Asp Thr His Gly Cys Gly Leu Arg Cys Ser Ala Leu Ser Asp
230 235 240
Pro Ser Gln Asp Leu Gln Phe Ile Val Ala Gly Asp Glu Cys Val
245 250 255
Tyr Leu Tyr Gln Pro Asp Glu Arg Gly Pro Cys Phe Ala Phe Glu
260 265 270
Gly His Lys Leu Ile Ala His Trp Phe Arg Gly Tyr Leu Ile Ile
275 280 285
Val Ser Arg Asp Arg Lys Val Ser Pro Lys Ser Glu Phe Thr Ser
290 295 300
Arg Asp Ser Gln Ser Ser Asp Lys Gln Ile Leu Asn Ile Tyr Asp
305 310 315
Leu Cys Asn Lys Phe Ile Ala Tyr Ser Thr Val Phe Glu Asp Val
320 325 330
Val Asp Val Leu Ala Glu Trp Gly Ser Leu Tyr Val Leu Thr Arg
335 340 345
Asp Gly Arg Val His Ala Leu Gln Glu Lys Asp Thr Gln Thr Lys
350 355 360
Leu Glu Met Leu Phe Lys Lys Asn Leu Phe Glu Met Ala Ile Asn
365 370 375
Leu Ala Lys Ser Gln His Leu Asp Ser Asp Gly Leu Ala Gln Ile

Phe Met Gln Tyr Gly Asp His Leu Tyr Ser Lys Gly Asn His Asp
380 385 390

Gly Ala Val Gln Gln Tyr Ile Arg Thr Ile Gly Lys Leu Glu Pro
395 400 405

Ser Tyr Val Ile Arg Lys Phe Leu Asp Ala Gln Arg Ile His Asn
410 415 420

Leu Thr Ala Tyr Leu Gln Thr Leu His Arg Gln Ser Leu Ala Asn
425 430 435

Ala Asp His Thr Thr Leu Leu Leu Asn Cys Tyr Thr Lys Leu Lys
440 445 450

Asp Ser Ser Lys Leu Glu Glu Phe Ile Lys Lys Lys Ser Glu Ser
455 460 465

Glu Val His Phe Asp Val Glu Thr Ala Ile Lys Val Leu Arg Gln
470 475 480

Ala Gly Tyr Tyr Ser His Ala Leu Tyr Leu Ala Glu Asn His Ala
485 490 495

His His Glu Trp Tyr Leu Lys Ile Gln Leu Glu Asp Ile Lys Asn
500 505 510

Tyr Gln Glu Ala Leu Arg Tyr Ile Gly Lys Leu Pro Phe Glu Gln
515 520 525

Ala Glu Ser Asn Met Lys Arg Tyr Gly Lys Ile Leu Met His His
530 535 540

Ile Pro Glu Gln Thr Thr Gln Leu Leu Lys Gly Leu Cys Thr Asp
545 550 555

Tyr Arg Pro Ser Leu Glu Gly Arg Ser Asp Arg Glu Ala Pro Gly
560 565 570

Cys Arg Ala Asn Ser Glu Glu Phe Ile Pro Ile Phe Ala Asn Asn
575 580 585

Pro Arg Glu Leu Lys Ala Phe Leu Glu His Met Ser Glu Val Gln
590 595 600

Pro Asp Ser Pro Gln Gly Ile Tyr Asp Thr Leu Leu Glu Leu Arg
605 610 615

Leu Gln Asn Trp Ala His Glu Lys Asp Pro Gln Val Lys Glu Lys
620 625 630

Leu His Ala Glu Ala Ile Ser Leu Leu Lys Ser Gly Arg Phe Cys
635 640 645

Asp Val Phe Asp Lys Ala Leu Val Leu Cys Gln Met His Asp Phe
650 655 660

Gln Asp Gly Val Leu Tyr Leu Tyr Glu Gln Gly Lys Leu Phe Gln
665 670 675

Gln Ile Met His Tyr His Met Gln His Glu Gln Tyr Arg Gln Val
680 685 690

Ile Ser Val Cys Glu Arg His Gly Glu Gln Asp Pro Ser Leu Trp
695 700 705

Glu Gln Ala Leu Ser Tyr Phe Ala Arg Lys Glu Glu Asp Cys Lys
710 715 720

Glu Tyr Val Ala Ala Val Leu Lys His Ile Glu Asn Lys Asn Leu
725 730 735

Met Pro Pro Leu Leu Val Val Gln Thr Leu Ala His Asn Ser Thr
740 745 750

Ala Thr Leu Ser Val Ile Arg Asp Tyr Leu Val Gln Lys Leu Gln
755 760 765

Lys Gln Ser Gln Gln Ile Ala Gln Asp Glu Leu Arg Val Arg Arg
770 775 780

Tyr Arg Glu Glu Thr Thr Arg Ile Arg Gln Glu Ile Gln Glu Leu
785 790 795

Lys Ala Ser Pro Lys Ile Phe Gln Lys Thr Lys Cys Ser Ile Cys
800 805 810

Asn Ser Ala Leu Glu Leu Pro Ser Val His Phe Leu Cys Gly His
815 820 825

Ser Phe His Gln His Cys Phe Glu Ser Tyr Ser Glu Ser Asp Ala
830 835 840

Asp Cys Pro Thr Cys Leu Pro Glu Asn Arg Lys Val Met Asp Met
845 850 855

Ile Arg Ala Gln Glu Gln Lys Arg Asp Leu His Asp Gln Phe Gln
860 865 870

875 880 885

His Gln Leu Arg Cys Ser Asn Asp Ser Phe Ser Val Ile Ala Asp
890 895 900
Tyr Phe Gly Arg Gly Val Phe Asn Lys Leu Thr Leu Leu Thr Asp
905 910 915
Pro Pro Thr Ala Arg Leu Thr Ser Ser Leu Glu Ala Gly Leu Gln
920 925 930
Arg Asp Leu Leu Met His Ser Arg Arg Gly Thr
935 940

<210> 22
<211> 336
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5870962CD1

<400> 22
Met Ser Lys Ser Val Pro Ala Phe Leu Gln Asp Glu Val Ser Gly
1 5 10 15
Ser Val Met Ser Val Tyr Ser Gly Asp Phe Gly Asn Leu Glu Val
20 25 30
Lys Gly Asn Ile Gln Phe Ala Ile Glu Tyr Val Glu Ser Leu Lys
35 40 45
Glu Leu His Val Phe Val Ala Gln Cys Lys Asp Leu Ala Ala Ala
50 55 60
Asp Val Lys Lys Gln Arg Ser Asp Pro Tyr Val Lys Ala Tyr Leu
65 70 75
Leu Pro Asp Lys Gly Lys Met Gly Lys Lys Lys Thr Leu Val Val
80 85 90
Lys Lys Thr Leu Asn Pro Val Tyr Asn Glu Ile Leu Arg Tyr Lys
95 100 105
Ile Glu Lys Gln Ile Leu Lys Thr Gln Lys Leu Asn Leu Ser Ile
110 115 120
Trp His Arg Asp Thr Phe Lys Arg Asn Ser Phe Leu Gly Glu Val
125 130 135
Glu Leu Asp Leu Glu Thr Trp Asp Trp Asp Asn Lys Gln Asn Lys
140 145 150
Gln Leu Arg Trp Tyr Pro Leu Lys Arg Lys Thr Ala Pro Val Ala
155 160 165
Leu Glu Ala Glu Asn Arg Gly Glu Met Lys Leu Ala Leu Gln Tyr
170 175 180
Val Pro Glu Pro Val Pro Gly Lys Lys Leu Pro Thr Thr Gly Glu
185 190 195
Val His Ile Trp Val Lys Glu Cys Leu Asp Leu Pro Leu Leu Arg
200 205 210
Gly Ser His Leu Asn Ser Phe Val Lys Cys Thr Ile Leu Pro Asp
215 220 225
Thr Ser Arg Lys Ser Arg Gln Lys Thr Arg Ala Val Gly Lys Thr
230 235 240
Thr Asn Pro Ile Phe Asn His Thr Met Val Tyr Asp Gly Phe Arg
245 250 255
Pro Glu Asp Leu Met Glu Ala Cys Val Glu Leu Thr Val Trp Asp
260 265 270
His Tyr Lys Leu Thr Asn Gln Phe Leu Gly Gly Leu Arg Ile Gly
275 280 285
Phe Gly Thr Gly Lys Ser Tyr Gly Thr Glu Val Asp Trp Met Asp
290 295 300
Ser Thr Ser Glu Glu Val Ala Leu Trp Glu Lys Met Val Asn Ser
305 310 315
Pro Asn Thr Trp Ile Glu Ala Thr Leu Pro Leu Arg Met Leu Leu
320 325 330
Ile Ala Lys Ile Ser Lys
335

<210> 23

<211> 163
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2818605CD1

<400> 23
Met Asp Asp Lys Glu Pro Lys Arg Trp Pro Thr Leu Arg Asp Arg
1 5 10 15
Leu Cys Ser Asp Gly Phe Leu Phe Pro Gln Tyr Pro Ile Lys Pro
20 25 30
Tyr His Leu Lys Gly Ile His Arg Ala Val Phe Tyr Arg Asp Leu
35 40 45
Glu Glu Leu Lys Phe Val Leu Leu Thr Arg Tyr Asp Ile Asn Lys
50 55 60
Arg Asp Arg Lys Glu Arg Thr Ala Leu His Leu Ala Cys Ala Thr
65 70 75
Gly Gln Pro Glu Met Val His Leu Leu Val Ser Arg Arg Cys Glu
80 85 90
Leu Asn Leu Cys Asp Arg Glu Asp Arg Thr Pro Leu Ile Lys Ala
95 100 105
Val Gln Leu Arg Gln Glu Ala Cys Ala Thr Leu Leu Leu Gln Asn
110 115 120
Gly Ala Asp Pro Asn Ile Thr Asp Val Phe Gly Arg Thr Ala Leu
125 130 135
His Tyr Ala Val Tyr Asn Glu Asp Thr Ser Met Ile Glu Lys Leu
140 145 150
Leu Ser His Gly Thr Asn Ile Glu Glu Cys Ser Lys Val
155 160

<210> 24
<211> 1983
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 381039CB1

<400> 24
gtcaggggag ccgggcgtgc ggaggcggct gatgaggcgg gaaggcggca gtggttgaag 60
gggtgattgc tgactagcgg ggagggagtg ggcagcgatg gtgatggcgg cgaagaaggg 120
cccgggcccg ggcggcgggg tcagcggggg caaggcggag gcggaggcgg cctcggaggt 180
gtggtgccgt cgggtgcggg agctgggtgg ctgcagccag gccgggaacc gccactgctt 240
cgagtgcgcc cagcgcgggg tcacctacgt ggatatcacc gtgggcagct tcgtgtgcac 300
cacctgctcc ggcctcctga gagggctgaa ccccccctcat cgtgtcaagt caatctccat 360
gacaactttc actgagcctg aagtagtatt cctgcaatcc cgtggaaatg aggtttgccg 420
gaagatttgg ttgggtctgt ttgatgctcg gacatcttta gtaccagatt ccagggatcc 480
tcagaaagtg aaggagtttc tccaggaaaa atatgagaag aagagatggt atgtcccccc 540
agaccaagtc aaggggccca cttataccaa aggcagtgcc tccacccctg tgcagggctc 600
catcccagaa gggaagcccc ttcggacact tctgggtgat cctgcaccgt ctctctcagt 660
tgctgcctcc acctcgagcc agcccgtcag tcagtctcac gctcggacat cccaggcccg 720
gagcactcag ccacctcccc actcctctgt caaaaaagcc agtactgacc tgctggctga 780
catcggtgga gacccctttg ctgcacccca gatggcacca gcttttgctg cattccctgc 840
ctttgggggc cagacacctt cccaaggagg ctttgccaac tttgatgcct ttagcagtgg 900
ccccagctct tctgtgtttg gaagcctccc tccagctggt caagcctcgt tccaggccca 960
gccaactcct gcagggagca gccaggggac tccatttggt gccactcccc tggcacccgc 1020
cagtcagcca aacagcctcg cagacgtggg cagcttcctg ggacccgggg tgcccgctgc 1080
aggtgttcct agcagcctct tcgggatggc tggccaggtc cccccgctcc agtctgtcac 1140
gatgggcggc ggcggcggca gcagcacagg gctggccttt ggagccttca ctaacccttt 1200
cacagctccc gccgcccagt ccccgctgcc ttccaccaac ccgttccagc ccaatggctt 1260
ggcgccaggg cccggctttg ggatgagcag tgctgggcct ggcttccccc aggcagtgcc 1320
acccactggg gcctttgcca gctccttccc agcaccgctg ttcccccccgc agaccccgct 1380
tgttcagcag cagaatggct cttccttcgg ggacttagga tcagccaagt tggggcagag 1440
gccactgagc cagccagctg ggatctccac caacccttc atgactggac cctcatcaag 1500

cccattcgcc tccaaacctc caaccaccaa ccccttcttg tagcactgtg tttttggggg 1560
gcctcttccc tgccttctgg ggcccctctg ctccctagag ctctggtgac cacttgcctg 1620
tgggcatttc tatgggcctt ggggatggtg gaggtgctaa tgctttgctt ggggcctaca 1680
ggtgaaaggt ggctgccctc agattccaca aagcctctct cccctccctc gcatcacccc 1740
cacccaggca ggaagcccag gaggagtgcg ggcagggcct gacctggagg agtgatggtt 1800
gagggcggag gatttttttc acatgatcag tcccctgtgg aacagctctt ccctgggcct 1860
agctcgccag cgctgtgctc ctcatggacg ggagcgcatg gggaaagtag gggaaagatg 1920
aggcacagtg ttgatggcgc agtgaccaga cagttctgag ccccagaagg cccatgtggc 1980
caa 1983

<210> 25
<211> 1559
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 383249CB1

<400> 25
cccagcgaga gatctgcagc taggctggct gcacttgctc cacgggtcag gggatcggag 60
ggggattgaa gaatgcgcca ttaaaaggaa agatcaagga gtaaaccaga agaagaagaa 120
aaagaggact tcaaagctgg gaaggatgag ttcttgcagc aacgtctgtg ggtccaggca 180
ggcacaggct gcagctgagg gtggttacca gcgctatgga gtccggtcct acctgcacca 240
gttttatgag gactgtacag cctcaatttg ggagtatgag gatgatttcc agatccaaag 300
atcacctaac aggtggagct cagtattctg gaaggttgga ctcatctcag gtacagtttt 360
tgtgatcctc ggattgactg ttctggcagt gggctttctt gtgccccca aaatcgaagc 420
atttggcgaa gccgattttg tggtggtcga cacacatgct gtccagttta acagtgctct 480
ggacatgtac aagctggcag gagctgttct cttctgcatt ggaggcacgt ccatggcagg 540
gtgcctgctg atgtcggtgt ttgtaaagag ctactccaaa gaagaaaaat tcctccagca 600
gaagtttaaa gaacgaatcg cagacatcaa agcccacacc cagccggtta caaaagctcc 660
agggccaggg gaaacaaaga ttccagtcac tttgtccagg gttcaaaatg tccagcctct 720
actggcaacc tgaaaccttt cccaccccag ttcttcttgt ctgatttatg cccgtggtta 780
aaaagagcag gccagttttc gagataaaga agatttggcg ttgactgccc tagggctgtg 840
ttcagctgtg ggcaatataa tgggtggact cacacttgct cagttcaggc agctctgctg 900
gagggcggtg ccatgcctaa gcctgtgcac atgcatccag ctgcttaaga tgggtgcttc 960
cttgtacccg gccaccagaa aacccctgga actcctcttt catagatcgt gactttgtgt 1020
tattttccgt gttgtttgaa agtgccgaac agtttagacc agctcaccaa tggctaatgt 1080
ggtccatttt attttctaat attggtttat ctaatgtttt ctgcgtggtg ctgtcttcca 1140
ttcctgtctc actatgtgta agattttgtc atatgtgttc atagaaacat ggaattctct 1200
gattttgagc cgaacatgag tttttagact gtatcttgac atgaggttcc tgacattgga 1260
tacaaagtta tcagcattca caaatctttt tgttttctcc attttttttc cctttgagat 1320
gaagtgctgt tcttttgcct aggttggagt acagcggtga aatcatagcg cactgcagcc 1380
ttgaactcct gggctcaagc gatcctccca cctcagcctc ccaagtagct gggactacag 1440
gcacctgcca ccacccccag ttaatatttt ttttaatttt ttttagaggc aggatcttgc 1500
tgtccaattt tttttaaaaa gctaaaattg ccaactttag aatcttgacc aaaaaaaaa 1559

<210> 26
<211> 1643
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 618769CB1

<400> 26
gaaatcatgc ccctcgtaga gcagcaggtc caagcagggc tgctggctat ttttccaaaa 60
agtgaggcag tttaaaaaaa aggcggagaa ctagaattat agaataatgg cacattttgt 120
gtatttgtaa aactaacggc ttgcatggtt cacaacccat ttcttatgcc tgtgttttcc 180
ttggcagcaa aatttctgtg gttcctccta ctccacctcc tgtcagcgag agccagtgca 240
gccgcagtcc tggcaggaag gtcagtgcac cagatattct gaaacctctc aatcaagagg 300
atcccaaatg ctctactaac cctattttga agcaacagaa tctcccatcc agtccggcac 360
ccagtaccat attctctgga ggttttagac acggaagttt aattagcatt gacagcacct 420
gtacagagat gggcaatttt gacaatgcta atgtcactgg agaaatagaa tttgccattc 480
attattgctt caaaacccat tctttagaaa tatgcatcaa ggcctgtaag aaccttgcct 540
atggagaaga aaagaagaaa aagtgcaatc cgtatgtgaa gacctacctg ttgcccgaca 600

gatcctccca gggaaagcgc aagactggag tccaaaggaa caccgtggac ccgacctttc 660
aggagacctt gaagtatcag gtggcccctg cccagctggt gacccggcag ctgcaggtct 720
cggtgtggca tctgggcacg ctggcccgga gagtgtttct tggagaagtg atcattcctc 780
tggccacgtg ggactttgaa gacagcacaa cacagtcctt ccgctggcat ccgctccggg 840
ccaaggcgga gaaatacgaa gacagcgttc ctcagagtaa tggagagctc acagtccggg 900
ctaagctggt tctcccttca cggcccagaa aactccaaga ggctcaagaa gggacagatc 960
agccatcact tcatggtcaa ctttgtttgg tagtgctagg agccaagaat ttacctgtgc 1020
ggccagatgg caccttgaac tcatttgtta agggctgtct cactctgcca gaccaacaaa 1080
aactgagact gaagtcgcca gtcctgagga agcaggcttg cccccagtgg aaacactcat 1140
ttgtcttcag tggcgtaacc ccagctcagc tgaggcagtc gagcttggag ttaactgtct 1200
gggatcaggc cctctttgga atgaatgacc gcttgcttgg aggaaccaga cttggttcaa 1260
agggagacac agctgttggc ggggatgcat gctcactatc gaagctccag tggcagaaag 1320
tcctttccag ccccaatcta tggacagaca tgactcttgt cctgcactga catgaaggcc 1380
tcaaggttcc aggttgcagc aggcgtgagg cactgtgcgt ctgcagaggg gctacgaacc 1440
aggtgcaggg tcccagctgg agaccccttt gaccttgagc agtctccatc tgcggccctg 1500
tcccatggct taaccgccta ttggtatctg tgtatattta cgttaaacac aattatgtta 1560
cctaagcctc tggtgggtta tctcctcttt gagatgtaga aaatggccag attttaataa 1620
acgttgttac ccaaaaaaaa aaa 1643

<210> 27
<211> 1613
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1234837CB1

<400> 27
cccgggccag gggcgggcgc cgccatgggt aacctgttcg gccgcaagaa gcagagccgc 60
gtcacggagc aggacaaggc catcctgcaa ctgaagcagc agcgggacaa gctgaggcag 120
taccagaaga ggatcgccca gcagctggag cgcgagcgcg ccctggcccg gcagctgctg 180
cgggacggca ggaaggaacg ggccaagctg ctgctcaaga agaagcgata ccaggagcag 240
ctcctggaca ggacggagaa ccagatcagc agcctggagg ccatggttca gagtattgag 300
ttcacccaga tcgaaatgaa agtgatggag gggctgcagt ttggaaatga gtgtctgaac 360
aagatgcacc aggtgatgtc cattgaagag gtggagagga tcctggacga gacgcaggag 420
gccgtggagt accagcggca aatagacgag ctcctggcag gaagcttcac tcaggaggat 480
gaagacgcca tcctggagga gctgagcgca atcactcagg aacaaataga gctgccagag 540
gttccctccg agccccttcc tgagaagatc ccagaaaacg tccctgtcaa ggccaggccc 600
aggcaggcgg agctggtggc agcttcgtaa cgtggcctcg tcttgtggga ctcacgggga 660
tgccccaggg actgtggccc acagagagtt tgggtcacgg ccagcccctg accgggttcc 720
ctggagccca gtgcgcacgg tgctgagcag agctgcagcc acgcaggcgc attgcaggag 780
gactccagag cgtctcctgg agaccttgag cctgaacgca ctcaggcgcc actggcctgc 840
tctcagtccg gattaactct cgaccgagcc cagcttctgc cggttgtggg ctccccggt 900
ggccgaggcc caggcccaac gcctctggtg ctgttccgct gcagtcccag ccccgcgtgg 960
ctcgcgctcg tctgtgagga agacacctcc agaccttggg gtccccgcgc ttcctcttgc 1020
tcctcgctgc tcccattagc tggtgcaggc ttccgttaag gggtccctcc cttggcctgg 1080
cttcccggcg cacctcagct tccctgctgg tgggggggatc cccaggagac cagcatgtgc 1140
tgaacctctc tgtgcctctg cctccgcacc ctagacaccc acctccagtt tgaaggtggc 1200
gggccagggg ctttcttgct gaattgacga ctccgagagc cctgactccc gccttgccac 1260
tcacggctct gtccactagg gctcagccct gctgagaaag gacctccgat gcttgggaga 1320
cgctgctccg gcaggtgcag ccccggaagt ttgtccatgg gggtccccgc ggctggggct 1380
catggaactg cgagacccgg gaccctcctg ccctgcgggt ccccgagcca ccagcagcca 1440
ggactggagg ctgtggggca tggcgtgact tctcgtgcac agggctggtt tggttatgag 1500
acgatctcgc tgggaccgcc cctgcccgtg gaaagccaca aggacaaagg gagcggccac 1560
ctcgaccccc agacaggccg gcccgtatta aagttggccc tgcacgccca aaa 1613

<210> 28
<211> 2810
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1607223CB1

<400> 28

tacaggttgg gcattgctgg gcactgtatg taggtcgcct gaccctatct cccatgggag 60
tgatgaccag tcccagaggc ctgtgtgact ctgaatcttc caacctcaca ggccattccc 120
tcccttctga aagcccaccc atgggcttct ggcccctgtc caaggcgtgg gagggaggcc 180
acggcctctg atgcacctgg gcgtcccctc ccatcccccc tccctgcaca ggtgctggaa 240
acatgcatga agagctgcgg caagcggttc cacgacgaag tgggcaagtt ccgctttctc 300
aacgagctca tcaaggtcgt gtctcccaag tatctgggct ctcggacatc ggagaaggtg 360
aagaacaaga tcttggagct cctctacagc tggacagtgg gcctgcccga ggaggtgaaa 420
atcgcagagg cctaccagat gctaaagaag caggggattg taaagtccga ccccaagctt 480
ccagatgaca ctacctttcc ccttcctcct ccacggccga agaatgtgat ctttgaagat 540
gaggagaaat ccaagatgct ggcccgcctg ctgaagagct cccatcccga agacctccgc 600
gcagccaata agctcatcaa agagatggtg caggaggacc agaagcggat ggagaagatc 660
tcgaagaggg tgaatgccat cgaggaggtg aacaacaatg tgaaactgct cacggagatg 720
gtgatgagcc acagccaggg cggcgcagca gctggcagca gcgaggacct catgaaggaa 780
ctgtaccagc gctgtgagcg gatgcggccc acgctcttcc gactggcgag tgacacagag 840
gacaatgatg aggccttagc ggagatcctg caggccaatg acaacctcac ccaggtgatc 900
aacctgtata agcagctggt gcggggtgag gaggtcaacg gtgatgccac agccggctcc 960
atccctggga gcacctcggc cctgctggat ctctcaggcc tggatctccc gcctgcgggc 1020
accacctacc cagctatgcc cacccgccct ggcgagcagg ccagccctga gcagcccagt 1080
gcctcagttt ccctgcttga cgacgagctc atgtctctgg gcctcagtga ccccacaccc 1140
ccttcaggcc caagcctgga tggtaccgga tggaacagct tccagtcgtc ggatgccact 1200
gagcccccag cccctgctct ggcccaggcc cccagtatgg aaagccgacc cccagcgcag 1260
acatccctgc cagcaagcag cggtctggac gacctagacc tcctggggaa gaccctcctg 1320
cagcagtcgc tgccccggga atcccagcaa gtgcggtggg agaagcagca gccaaccccc 1380
cggctcacac tccgggacct gcagaataag agcagcagct gcagctcccc cagctccagc 1440
gccaccagcc ttctccacac cgtgtcccca gagccccccа ggcctccgca gcagcccgta 1500
ccaaccgagc tctcactggc cagcatcact gtgcccctgg agtccatcaa acccagcaac 1560
atcctgcccg tgactgtgta tgaccagcac ggcttccgca tcctcttcca ttttgcccgg 1620
gacccactgc cagggcgctc cgacgtgctg gtggtggtgg tttccatgct gagcaccgcc 1680
ccccagccca tccgcaacat cgtgttccag tcagctgtcc ccaaggttat gaaggtgaag 1740
ctgcagccac cctcgggcac ggagctgcca gcttttaacc ccatcgtcca cccctcagca 1800
atcacccagg tcctgctgct tgccaacccc cagaaggaga aggttcgcct ccgctacaag 1860
ctcaccttca ccatgggtga ccagacctac aacgagatgg gggatgtgga ccagttcccc 1920
ccacctgaaa cctggggtag cctctagaac agaggggctg gggagaggaa ggggcagagg 1980
gaccggtcac tgtccagcct ggagggaggc attggtggcc aaggacaccc tttgttgccc 2040
atggccattc acccccaggc ctggtgcttc tccccacacc cctgtaggcc tcaagtgact 2100
cttccccctc ctgctccggc cccgcccctg ctgagccaaa cccagtagga ggctgggcct 2160
gggtttgtgc cgctggggtc tccatcaccg ggacctggag agggaggggc tgtgtagcct 2220
tggaagaact tgggtcatgg ggaggaagca cagctgttgg ggaagggcca ggacctcagg 2280
cccagcccca accccagctg gggtggggtc ttccccacct gtctcttatg ccttatggga 2340
aggcccagcc ataactcggg ggccatgctg gagctgggga ccagcttagg cctcctccat 2400
aggaacccag tgactggggg gtgacgccta caccccagc tatttgcact ctggtgtgtg 2460
gtttgactct gcttttcttc cggattggcc ctgtggtcac agcctcaggg ggccaggctg 2520
ggggaacctc acctggcccg tactcctggg ggtttccctt tgccattggg ccccctgagg 2580.
gactgtgggg gctcaagggt aatgccagag gcccatggcc ccagcgaggg gctgtggggc 2640
acctagagtt ctcggtgtgt ctccttcatt cattggcctc tgctggggcc tcctatgggt 2700
gtcttacgtc tgtccatcca tctgtccgtg gtcagaagtg gggtcagtgt gtgagtgaga 2760
gcaggagtat ttatgaaaat aaaacgtcgt ttttcctgga aaaaaaaaaa 2810

<210> 29
<211> 2321
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1621554CB1

<400> 29
tcccaagtga gtcgaggggg gatcccgact ccagtccggg gccttggcca gcggagccgc 60
ggtattcgga agcgggaatc ccactcagag cccgggcctg taggggcggg gcgtcccggg 120
cacccgggat tggggcgtct cccgtcgtgc accggggcac cggcgactca cccggaagga 180
gaagccgtga tctggctata tggtggggcg cgggcggtgt cgctgtgggg agctggtgct 240
gttctcagat gtttccttcc aatgggcttt tggtgtagga tgtcggagaa ccaagaacag 300
gaggaggtga ttacagtgcg tgttcaggac ccccgagtgc agaatgaggg ctcctggaac 360
tcttatgtgg attataagat attcctccat accaacagca aagcctttac tgccaagact 420
tcctgtgtgc ggcgccgcta ccgtgagttc gtgtggctga gaaagcagct acagagaaat 480
gctggtttgg tgcctgttcc tgaacttcct gggaagtcaa ccttcttcgg cacctcagat 540

gagttcattg agaagcgacg acaaggtctg cagcacttcc ttgaaaaggt cctgcagagt 600
gtggttctcc tgtcagacag ccagttgcac ctattcctgc aaagccagct ctcggtgcct 660
gagatagaag cctgtgtcca gggccgaagt accatgactg tgtctgatgc cattcttcga 720
tatgctatgt caaactgtgg ctgggcccag gaagagaggc agagctcttc tcacctggct 780
aaaggagacc agcctaagag ttgctgcttt cttccaagat cgggtaggag gagctctccc 840
tcaccgcctc ccagtgaaga aaaggaccat ttagaagtgt gggctccagt tgttgactct 900
gaggttcctt ccttggaaag tcccactctc ccacccctct cctcaccatt atgctgtgat 960
tttggaagac ccaaagaggg aacctccact cttcagtctg tgaggagggc tgtgggagga 1020
gatcatgctg tgcctttgga ccctggtcag ttagaaacag ttttggaaaa gtgagctctc 1080
ggttctgctc tgagatggtc agagaagatg cgggccagga gacttactca ggtgggactg 1140
ggcacagggc aggtatgtgg gaggctgggc tgcttagtgt cttctagtca cctctgcttg 1200
ggctgattga cagaggtcag tcattacagc cccttatgcc tcttccatgg gaacaaatac 1260
tgtgcagatg tttgtaagtt aaacataaga cacaggggct gttgcttttg aacagaaccc 1320
tatattactc tcctgggatc tgagtttctg caggtcattt gtatgtagga ccaggagtat 1380
ctcctcaggt gaccagtttt ggggacccgt atgtggcaaa ttctaagctg ccatattgaa 1440
catcatccca ctgggagtgg ttatgttgta tccccatctt ggctggcttc agtttttgct 1500
gtagccctag agcactttgt ttgtgggagg ctggcctctt gcctacctcc ttgcatggac 1560
aggggatga atatttactt tcccacctcc ttgctttttc tttcactgat accactgaat 1620
ggaactggtg ctgtgactcc tgctgctggg gatttatgtc ccgagacctt agcctggctg 1680
agtggagcct gagacctgca caacagctca tggtcatgca tgagagagaa gtggctggcc 1740
acagccagag ggaacagtaa cagcccaggg gcctttattt tgggaaaggc tgtccggggc 1800
tgttactgtc tcttctggtt ataaagcaga catgtggcca tcttttccgc agggttagag 1860
tgggctcctt tctttttgga atccttttct tctcctttgg tagcagctcc ctgcctccag 1920
ggcttccgcc accagcgtct ctgctgtgtt gcgcagtgca gtggggtgca agggctttgt 1980
ttctgcctgc ctgaaagaga gggctctggg gatggagatg agaaacaaca cgctctcctt 2040
cagacaatga ggcattctgt cctcctgctg ccattcttca tctccactga gagccagagc 2100
tggtaggagc cgagtgccac aggcattctg cattgctcta ctcttaggtt tgtgtgtgtg 2160
atccttcccc tccctgtcgc ccactcctcc ctcctctggc tatcctaccc tgtctgtggg 2220
ctcttttact accagcctat gctgtgggac tgtcatggca tttagttcag agtggagggg 2280
ctttggcctg aaataaaatg caagtattta aaaaaaaaaa a 2321

<210> 30
<211> 3581
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1751553CB1

<400> 30
gcggccgcgc gggcggccag cgggcgttgg gacgtggaga cccggagagg gcggcgggac 60
ccgggcgggg aaaggcgcgg cgtggcttgg ctcaggtgcg cttctcccac ctggcagctc 120
gctcagggct gtgggggcgg cctgtgaggg cgcccgccgt tgcggcgtgg gagattaata 180
ttaagattgg aagtttgtgt cttttgctgg atattggaaa ttgaatgtaa tggcaacaga 240
atttataaag agttgctgtg gaggatgttt ctatggtgaa acagaaaaac acaacttttc 300
tgtggaaaga gattttaaag cagcagtccc aaatagtcaa aatgctacta tctctgtacc 360
tccattgact tctgtttctg taaagcctca gcttggctgt actgaggatt atttgctttc 420
caaattacca tctgatggca aagaagtacc atttgtggtg cccaagttta agttatctta 480
cattcaaccc aggacacaag aaactccttc acatctggaa gaacttgaag gatctgccag 540
agcatctttt ggagatcgaa aggtagaact ttccagttca tcccagcacg gacctagcta 600
tgatgtgtat aacccattct atatgtatca gcacatttca cctgatttga gtcgacgctt 660
tcctccccgt tcagaagtga cgagactgta tggatcggtt tgtgatttaa ggacgaacaa 720
acttcccggt tcccctgggc taagcaaatc tatgtttgat cttacaaact catctcagcg 780
attcatccag agacatgatt cattgtccag tgtacccagt agttcttctt caaggaaaaa 840
ttctcagggg agtaacagaa gcctggatac aattactcta tcaggagatg aaagggactt 900
tgggagactg aatgtgaaat tgttttataa ttcttcagta gaacagatct ggatcacagt 960
tttacagtgc agagatttaa gttggccctc tagttatgga gacactccta ctgtttctat 1020
aaaaggaata cttacattgc ccaaaccagt gcatttcaaa tcttcagcca aggaaggttc 1080
caacgctatt gaatttatgg aaacgtttgt atttgctatt aaacttcaaa atctacaaac 1140
tgtaagactt gtatttaaga ttcaaaccca gactcccagg aagaaaacca ttggagaatg 1200
ctcaatgtca ctcagaaccc ttagcacaca ggaaatggat tactctttgg atataacacc 1260
accttcaaaa atttctgttt gccatgcaga acttgaattg gggacttgtt ttcaagcagt 1320
aaatagcaga attcagttac aaattcttga ggcacggtac cttccaagct catcaacacc 1380
tctgactttg agttttttcg tgaaggtggg aatgtttagc tcgggagagt tgatttataa 1440
gaaaaagaca cgcttactga aggcctccaa tggaagagtc aagtggggag agactatgat 1500
ttttccactt atacagagtg aaaaagaaat tgtttttctc attaagcttt acagtcgaag 1560

ctctgtaaga agaaaacact ttgtgggcca gatttggata agtgaagaca gtaataacat 1620
tgaagcagtg aaccagtgga aagagacagt aataaatcca gaaaaggttg ttatcaggtg 1680
gcacaaatta aatccatctt gaagacttca cacattaatt tggtgaagaa cttgacattc 1740
ttttagaaga cttatgattt caatttgcta ccaatgagaa gaggcaaatc aacaaatttg 1800
tcaatttatg ggggctataa ttatggtata taatgtatct gatagaaaat ttgataagaa 1860
aatgtaatga attttatcag atatccaaag taaaggaaat gttttaaaac tgcaacaaga 1920
gacacagaca gtaaaatcaa agtattatta ggatgactaa ataaattata aagtctgtga 1980
gaatatcaac catagatagt tctttctata ttatgttttt gcttttgtat tttaagcttt 2040
acttagatat tcaaaacctg gtatatcaag tctctgttag tactattggc atttagaaga 2100
ctttaccatt atttcagtgc taggcattat tgattaggtc ttggctccac tgtttacctc 2160
ttgctatgta ttttctcccg gtaaaaatga attgaaccat ttcaactatt ttctatattt 2220
ggagaaagtt tgtgccctgt gttttataat ttttttaccc ataagacatc acattatccc 2280
tttgtaagct acttatctcc aaaaaacttc agaaatagaa aactacattt tggcaggaat 2340
aattgaaaac accagaaggt tgaagtttaa ttggaaaccc agaatataca tactttgctg 2400
ttttcttccc tcaaatattt tactatttgt tttatttgga gttaaaataa gagtatcatc 2460
catatggtcc atcctaattc acagaattaa atgagcttaa atagaaaatt cagtatttta 2520
tgataatcac ttcgttttta gtttttaaaa tttagattat tctataattt accgtgtttg 2580
agtatttttct catttttttc ataaccatac ctgattatac tgtgtaacaa atattttcta 2640
ttgcagtttt ctttccagta cttattagaa ctcagtattt ggaaataatt tcagcttaat 2700
tgaccataag aactgtggcc aaaaagaaca gtttttttgga gaggcagatg acattatacc 2760
tgattttaga aaatctcact ttatttttgc taataagtag actaagtgct ctgtgttctc 2820
agtcttccct ttttttctgc ccccattctt actttgtccc aggcatgcag agaaagatgg 2880
tgatatttta ggccaggagt ataccttgct ataacctaag catgccttct ttattccagc 2940
tcctatgttc tgtgtatatc attaacattt tcccaaataa acacttaatt ctctttttccc 3000
taggtgccat ctcctcaagc tacaaaatgt ccacatctta tatccccttt gcttctactg 3060
ctctgatttt gtggtaccag tactctctgc cactgaacat tttgaaatat ttttgtttta 3120
gatttgcaaa aaatgacata taggtcagta ctcacatgga tttttaagat aaatcacctg 3180
tgtgataata ttttgaatct gagacgaata caacttttaa aaattgtttt taaaaataga 3240
ctttttttttt tagagcagtt gtaggttaac agaaaaattg agaggaagat agagatttcc 3300
tttctcccct gacaaagccc tcaacagcct cccaggctat cagtatcctg caccacaatg 3360
gtacatttgt tacaatcaat gaacctactc tgaaacatca ttatcatcca aagttcatgg 3420
tttacattag agtcccctct tggtgttata catgctagag gacaaatata tgatgatatg 3480
tatgcatcat tataatatag tatagtttcg ctgccctaaa catcctctgc aaataaaact 3540
attttaatga aacctaaaga agaaaatgta tagaaatacg t 3581

<210> 31
<211> 1014
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1832403CB1

<400> 31
cccacgcgtc cgccggcggc cctcgcggca ggtttcgggc ttcaggacaa ttcgtgatgg 60
cgggggctgg ttccgccgct gtatcggggg cagggacccc ggtggcgggg cccacaggcc 120
gcgacctttt cgccgaaggg ctgctggagt tcctgcgacc cgctgtgcag cagctcgact 180
ctcacgtaca cgccgtcaga gagagccagg tagagctccg ggaacaaatt gacaacctag 240
ccacagaact gtgccgcata aatgaggatc agaaggtggc cctggatctt gacccctatg 300
ttaagaagct acttaatgcc cggcgacgcg ttgtcttggt taacaacatt ctacagaatg 360
ctcaggaacg actgagacgg ctaaaccaca gtgttgccaa ggaaacagcc cgcaggagag 420
caatgctgga ttcgggaatt tacccccctg gctccccagg caaataacag atgagcctat 480
ggactcagta gcacaagtac tgttccccag ctgccttgtt tcaacagaca tgcaaagatc 540
ctaggagaca gtccccatag accttcagac attaaaaagg gagccgtaca gtttgtttga 600
agcacttcgt cttacccatt tatgtagggg ccccaggaaa cctacacaca gccagaatga 660
ggttcccaaa ggacttacat taattatggc tcttgcttcc tttcacaaat gagctgaggc 720
ctctactttt ttttttaagc tgcatacttg aggcttacct tcttcaggac tagttaacca 780
gaggggcttc ctttgtatgt tacatgcctg gttacatggg cctggacagc atgtcctcta 840
cctgtgactt ctcattttcc tgtttacact ggggatttgg agggggcagg caaagtcaaa 900
gtgaatgacc tctgtccacc cacttttta ttgcactggc ttgaatacag tagcagtgtt 960
gatagaatca ttttattcaa taaatactta aaatgataaa aaaaaaaaa aagg 1014

<210> 32
<211> 4195
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1971747CB1

<400> 32
gcaagactag gcaacctcca gccagtccct gggtcgggcg gatcctccca gaggtggcac 60
aatggagcga tctccaggag agggccccag ccccagcccc atggaccagc cctctgctcc 120
ctccgacccc actgaccagc cccccgctgc tcacgcaaag ccagacccag gttctggggg 180
ccaacctgct ggccctggcg cggcgggtga ggccctggcg gtgctgactt cattcgggag 240
gcggttgctg gtgctgatac ctgtgtattt ggccggggca gtgggactca gcgtgggttt 300
cgtgctcttc ggcctcgccc tctacctggg ctggcgccgg gtccgcgacg agaaagaacg 360
gagccttcga gcagcgaggc agctactgga cgacgaggag cagctcactg cgaaaactct 420
ctatatgagt catcgagagc tacctgcctg ggtcagcttc ccagacgtgg aaaaggctga 480
atggctcaat aagattgtgg cccaggtctg gcccttcctg ggccagtata tggagaagct 540
tctggctgaa actgtggctc cggctgttag gggatctaac ccccatctgc aaacatttac 600
atttacacga gtggaactgg gtgaaaagcc attgcgcatc attggagtca aggttcaccc 660
aggtcagaga aaagagcaga tcctgctgga cttgaacatc agctatgtag gtgatgtgca 720
gattgatgtg gaagtgaaga aatatttttg caaagcagga gtcaagggca tgcagctaca 780
tggcgttttg cgggtgatac tggagccact cattggggac cttcccttcg tgggggctgt 840
gtcaatgttc ttcatccgac gcccgacctt agacatcaac tggacaggga tgaccaacct 900
gctggatatc ccaggactta gctcactctc tgacaccatg atcatggact ccattgctgc 960
cttcctcgtg ttgcccaacc gattactggt gccccttgtg cctgaccttc aagatgtggc 1020
tcagttgcgt tcccctctgc ccaggggcat tattcgaatt cacctgctgg ctgctcgagg 1080
gctgagttcc aaggacaaat atgtgaaggg cctgattgag ggcaagtcag acccatatgc 1140
acttgtgcgt ttgggtaccc agacattctg cagtcgtgtc attgatgaag aactcaaccc 1200
acagtgggga gagacttatg aggtgatggt acacgaggtc ccagggcagg agattgaagt 1260
ggaggtgttc gacaaggatc cagataaaga tgactttctg ggcagaatga agctggatgt 1320
agggaaggtg ttacaggcta gcgttctgga tgattggttc cctctacaag gtgggcaagg 1380
ccaagttcac ttgaggctag aatggctgtc acttttgtca gatgcagaga aactggagca 1440
ggttctacag tggaattggg gagtctcctc tcgaccagat cccccgtcag ctgccatctt 1500
agttgtctac ctggatcggg cccaggatct tcctctgaag aaggggaaca aggaacccaa 1560
ccctatggta caactgtcaa ttcaggatgt gactcaggag agcaaggctg tctacagtac 1620
caactgccca gtgtgggagg aagcgttccg gttcttccta caagaccctc aaagccagga 1680
gctcgatgtg caagtgaagg atgattccag ggccctgact ttaggagcac tgacgctgcc 1740
tctggcccgc ctgctgactg ccccagaact catcctggac cagtggttcc agctcagcag 1800
ctctggtcca aactccagac tctatatgaa actagtcatg aggatcctgt acttggattc 1860
atcagaaata tgcttcccca cggtgcctgg ttgtcctggt gcttgggacg tggacagtga 1920
gaatccccag agaggcagca gtgtggatgc cccacctcga ccctgtcaca cgactcctga 1980
tagccagttt gggactgagc atgtgcttcg gatccatgta ttagaggccc aggacctgat 2040
tgccaaagac cgtttcttgg ggggactggt gaagggcaag tcagaccccc atgtcaaact 2100
aaagttggca ggacgaagct tccggagcca tgttgttcgg gaagatctca atccccgctg 2160
gaatgaggtt tttgaggtga tcgtcacatc agttccaggc caagagctag aggttgaagt 2220
ctttgacaag gacttggaca aggatgattt tctgggcagg tgtaaagtgc gtctcaccac 2280
agtcttaaac agtggcttcc ttgatgagtg gctgaccctg gaggatgtcc catctggccg 2340
cctgcacttg cgcctggagc gtctcacccc ccgtcccact gctgctgagt tagaggaggt 2400
gctgcaggtg aatagtttga tccagactca gaagagtgcg gagctggctg cggccctgct 2460
atccatctat atggagcggg cagaggacct cccgctgcga aaaggcacca agcacctcag 2520
cccttatgct actctcactg tgggagatag ttctcataaa accaagacta tttcgcaaac 2580
ttcagcccct gtctgggatg agagtgcctc ctttctcatc aggaaaccac acactgagag 2640
cctagagttg caggttcggg gtgagggcac tggcgtgctg ggctcattat ccctgccct 2700
ctcagagctc ctcgtggctg accagctctg cttggaccgc tggtttacac tcagcagtgg 2760
tcaggggcag gtgctactga gagcacagct agggatcctg gtgtcccagc actcgggagt 2820
ggaagctcat agccacagct acagccacag ctcctcatcg ctgagtgaag aaccagagct 2880
ctcgggggga ccccctcaca tcacctcctc agccccagag ctccggcagc gcctaacaca 2940
tgttgacagt ccccttgagg ctccagccgg gcctctgggc caggtgaaac tgactctgtg 3000
gtactacagt gaagaacgaa agctggtcag cattgttcat ggttgccggt cccttcgaca 3060
gaatggacgt gatcctcctg atccctatgt gtcactgttg ctactgccag acaagaaccg 3120
aggcaccaag aggaggacct cacagaagaa gaggaccctg agtcctgaat ttaatgaacg 3180
gtttgagtgg gaactccccc tggatgaggc ccagagacga aagctggatg tctctgtcaa 3240
gtctaattcc tccttcatgt caagagagcg tgagctgctg gggaaggtgc agctggacct 3300
agctgagaca gacctttccc agggtgtagc ccggtggtat gacctgatgg acaacaagga 3360
caagggcagc tcctaggagc tggcgagtcc cagcctgact gctctgtctt cctgccttcg 3420
tctcgctcca tcaccgcctc aatgtgatga gcctaaagct agggtccaag ggcagagcct 3480
gtgcccttca gcccttcac ctaacaggcc catattcggg cctttgcctg accaaagaga 3540
agaaccgtat gttcccttta ctgcacggcc tttatccttc tgggcccctg gggcggggac 3600
ctgagctggc tgtttcctgc tttgcctgca cattgttctc ccttcctccc aactcctcag 3660
ggccttctgt atctgtgcct ggccagtggc agcactagca gtggtattag cttatgccaa 3720

atacagcttt ggaaggatct ttttttcttt aactagatgg tcaccttctt ccctaccaca 3780
catgggtggg aaggtggaca ggctaacctc tccagctgtg agcctcttag actactgcat 3840
gtagcaaatg ttcagcagct caggccccca tgtccagttc tgtccccact gtcctcaacc 3900
ctgtcctgaa aattctactg ctttgatggc tggggccagt ctcttgtcac tttggaaact 3960
gaggacgcgt ggattctact caagcctcca agtagtggca tatcagtctt ggagctccta 4020
gctggtgata cggagagggc tttggaggac ttgggacagc agggccaatt tttttgccca 4080
agtgcctagg ctgctaactc actgactaga acttaatctg gtactttaca gttttgcacc 4140
aactctgcca agccactgga tcttacatta aacatcatac tcaaaaaaaa aaaaa 4195

<210> 33
<211> 3343
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2285348CB1

<400> 33
gaaaaagaca aagatgatct ggggcctgac agattctcaa cactcacaga tgatcccagc 60
cctagactca gtgcacaagc tcaggtggct gaggatattc tggacaaata caggaatgcc 120
attaaacgga ccagccccag tgatggagca atggcaaact atgaaagtac agaggttatg 180
ggtgatggtg aaagtgcaca tgattctccc cgtgacgaag cactgcagaa catctcggct 240
gatgatctcc cagactctgc aagccaagca gcccacccgc aggattcagc tttctcttac 300
agagatgcaa aaaagaaact gaggcttgct ctttgctctg cggactctgt tgccttccca 360
gtgctgaccc attcaacaag gaatggttta ccagaccaca cagacccaga agacaatgaa 420
attgtatgct tcttaaaagt tcaaatagct gaagcaatta atttacaaga taagaatcta 480
atggctcaac ttcaagaaac aatgcgctgt gtgtgccgtt ttgataatag gacttgtagg 540
aaactgctgg cttcgattgc tgaggactac agaaaaagag ccccatatat tgcttatctc 600
actcgttgtc gacaaggact acagaccaca caggctcacc tggaaaggct attgcaaaga 660
gttttgcggg acaaagaagt ggccaatcga tactttacca ctgtctgtgt gagattactg 720
cttgagagca aagaaaagaa gatcagggaa ttcattcaag actttcagaa actcaccgca 780
gctgacgata aaactgctca ggtagaagat tttctgcagt ttctttatgg tgcaatggcc 840
caggatgtca tatggcaaaa cgcgagtgaa gaacagcttc aagatgcaca gctggccatt 900
gagcgaagcg tgatgaaccg gattttcaag ctcgccttct accctaatca agatggggac 960
atacttcgcg accaggttct tcatgaacat atccagagat tgtctaaagt agtgactgca 1020
aatcacagag ctcttcagat accagaggtt tatcttcgag aagcaccatg gccatctgca 1080
caatcagaaa tcaggacaat aagtgcttat aaaacccccc gggacaaagt gcagtgcatc 1140
ctgagaatgt gctctacgat tatgaacctc ctgagcctgg ccaatgagga ctctgtccct 1200
ggagcggatg actttgttcc tgtgttggtg tttgtgttga taaaggcaaa tccaccctgt 1260
ttgctgtcta ctgtgcagta tatcagtagc ttttatgcta gctgtctgtc tggagaggag 1320
tcctattggt ggatgcagtt cacagcagca gtagaattca ttaaaaccat cgatgaccga 1380
aagtgaccaa gaccaaggcc caccaaggca gcagactgtt aatcagacaa acagatctct 1440
gagaaggtgc atcagctgct ttgaaggctg aagattgttt tgtatgatac tgcacagcat 1500
caggcatttt aaagcagatc tttactaaac aggttaatga gctaacaagc aggttctctc 1560
gtctttgggc tctttccttt ctgagttgca tattctattt tcttgtcccc aagtagagac 1620
tagtactaca aaaagggacc acatttttca agtatttta agtataaaaa acaaaacaaa 1680
aatctcttag gaaatgtcta gacctccatt cttggattcc ctttctttcc ttttatttta 1740
aaaaagaaca gtacccctct tttaagatgc tgtcttacat taatgagcat ctaatggaaa 1800
gaaggtatga gttgcactga ggattagaat agtggtgcgt tagtggcatt atctataaat 1860
acactcacct aaattgaaag ctaagaagga aatgtaaata taatatatat ttatatttga 1920
tgtaatatgg acatctgcag attctaataa acaaggacta ttgctgatag taggctgtga 1980
catactgtct tgtgaaatgg tttccttgac aaaatttaag ctgagcttaa aagcaaaaaa 2040
caaaaagtac acagaaatat ttattaaaat gtaatacagt ttattgaact ttctaggtat 2100
ggagtttgat ggacagggct gcctttaatg agtgtgaagg tcactaagtc acttagacat 2160
ctcaccgtgg aagtttgtga gcctgcatta ggagatagac tgattaccat acatgacata 2220
aaaaggaaca gtggatagct catactttat ggtggttctt ctcctccgaa ataatatact 2280
gcagaaatcc cagacagagc tccttacaaa cctttaattg taatatattt ttgatgatta 2340
ttcacattga atgcacagac caagaattca gtgaatgtca ttttttaaaa aactaatttg 2400
tattgtctgc tctagtgata caagttttac tagtgataaa ctattttaat caaccatact 2460
attcttatgg aaaaaaatat ctattttggc aggtttctgt gcctttattt ccctcttctg 2520
aaaaaaagtc tgtgttttca tagtttggtt tgcattgtat atcaataatt aatcaggaat 2580
gggttttggt gcctgaaaaa ttggccatgg aggcacacca aagcttcaag cacaagtctt 2640
gtacatgggc catcactgtc tggtttcact tcgtgtgttt cctaaacaca tttagctgct 2700
tttttaacaa actcagcccc atacttgagt cccttgttgt tgggagcatt tccaggcatc 2760
ttttaaggga actgtgacaa acagcctcgg gcagatgaac acggaggctc tctgttgtct 2820
gtctctgaga tctttgtgtc tgggaatgcc taaagatttt attttttttt ctttgttttt 2880

attttatttt attttatttt tttgagacag agtctcaccc tgttgcccag gctggagtgc 2940
aatggtgcga tcttggctca ctgcaacctc cacctcccag ttcaagtgat tcccctgcct 3000
cagcctcgcg agtagctagg actacaggcg catgtcacca agcccggcta atttttgtat 3060
ttttagtaga aacggggttt caccatgttg gccaggatga tcctcaatct cctgacctcg 3120
tgatccaccc gccttggcct cccaaagtgc gggattacaa gcgtgaacca ccctgcccag 3180
ccagaaacta gattttcttt atgcctccac cccttcttat tcatttactt tacaatacca 3240
aaataacaaa ttgcatagga gtgtgggatg tggtttctgc cttctagaga gagatcagaa 3300
ttaaactagt actatgaaat gtccttttga atgttaggtc aag 3343

<210> 34
<211> 2456
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2374186CB1

<400> 34
tgagtcagca aactccgcgg cccgcaagcc cggctcggcc cggccctgct ctgttctgcc 60
cggaggagcc gcccattgat cgtgtcctgt gctgaagatg tttccggaac aacagaaaga 120
ggaatttgta agtgtctggg ttcgagatcc taggattcag aaggaggact tctggcattc 180
ttacattgac tatgagatat gtattcatac taatagcatg tgttttacaa tgaaaacatc 240
ctgtgtacga agaagatata gagaattcgt gtggctgagg cagagactcc aaagtaatgc 300
gttgctggta caactgccag aacttccatc taaaaacctg tttttcaaca tgaacaatcg 360
ccagcacgtg gatcagcgtc gccagggtct ggaagatttc ctcagaaaag tcctacagaa 420
tgcacttttg ctttcagata gcagccttca cctcttctta cagagccatc tgaattcaga 480
agacattgag gcgtgtgttt ctgggcagac taagtactct gtggaagaag caattcacaa 540
gtttgcctta atgaatagac gtttccctga agaagatgaa gaaggaaaaa aagaaaatga 600
tatagattat gattcagaaa gttcatcctc tgggcttgga cacagtagtg atgacagcag 660
ttcacatgga tgtaaagtaa atacagctcc gcaggaatcc tgaaaaataa ttctaatgtt 720
actatcttag gaatagcaaa ttatgtccag tcatagagaa gaaagcttca taataataca 780
ttcttaccta aagctcactg tcatgatgtt aggtatttaa attcttaaag atgttgggtt 840
gtttattagt ggtatttta tgttgtctta ttttaggtaa gcttctgtgt aaagctaaaa 900
atcctgtgaa tacaatacta tcctttacag gcagacatta ttggtaaaca agatcttgcc 960
ctccaatgaa atgacttaca tgttttaaaa aaccgagttg gttttattga atttaaaaag 1020
ataggtaact aagtagcatt taaaatcaag atagagcatt ccttcttgta tcagtggggc 1080
agtgttacca taaacacggt gtatatgttg ttaaacccta tgaagagtaa cagtgtagac 1140
cagactgcct ctctcagata tgtgcctgat attttgtgga tacctcccct gcactggcaa 1200
aacactatgc ttttgggtgt tagactgaaa tattttaaga gtatttaacc tttccagtat 1260
tctgtttcac gcttagatgg aaatgtatct tatgaataga gacatattaa aataatgttt 1320
acatcttaga aaaaacatag atagtgctag taatattact tataactgta atatatagat 1380
tcagaaatac attttcatta tccaaaatca gcttcaacaa atggtttctg gagacaaata 1440
atttgttttc attatcattg tataatcagg ttaatgattt atttttgac taaatgtgca 1500
atttcttatc actagataac tttcagtatc agtggtggtt acttattact taaatcagag 1560
gaaggatttt ataaagatta ataaatttaa ttttaccaat aaatattccc ataatttaga 1620
aaaggatgtc gacttgctaa tttcagaaat aattattcat ttttaaaaag ccccttttaa 1680
agcatctact tgaagattgg tataattttc ataaaatgtc tttttttta gtgtcccaaa 1740
gatatcttag ataaactatt ttgaagttca gatttcagat gaggcaacat tttcttgaga 1800
taattaccca agtttcatcc atgttgaatg gtacaaaata tttctgtgaa actaacagga 1860
agatattttc agataactag gataacttgt tgctttgtta cccagcctaa ttgaagagtg 1920
gcagaggcta ctacaaaaag caacctttc attttcacta agagtttaaa agctattgta 1980
ttattaaaaa gtctttacaa tgcttgtttc aaagaaccaa cagaaaaaaa agctaagaaa 2040
actgagaact aacattaaaa aaattaaatt tagaataaga atgatttctt taatttgtcc 2100
ttttttttctt tggtctaaaa cattattaaa tttttgtaaa tattttgatt taatgtgtct 2160
tagatcctca ttattttaat acaggaaaag aaaagattta gtaatttctt accatgctaa 2220
tatgtaaagt tcatgccatc caggcattta agagcgatcc tcatcccttc agcaatatgt 2280
atttgagttc acactatttc tgttttacag cagttttgaa aaacacatac tatgccacca 2340
attgtcatat tatttttaga tgatgtaaca tagccatcaa aattaatatt atgtaatgcc 2400
taatacttag tatgtaaatg tcacgagatc atttttacat taaacgtgaa aaaaaa 2456

<210> 35
<211> 2004
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 2476232CB1

<400> 35
cgcggacagc ggtggaggcg gatttcctgg gcccggccct ctggcgctac catggcgttt 60
ggcaagagtc accgggatcc ctacgcgacc tccgtgggcc acctcataga aaaggctaca 120
tttgctggag ttcagactga agattggggc cagttcatgc acatctgtga cataattaac 180
actacccagg atgggccaaa agatgcagtg aaagctttga agaaaaggat ttccaaaaac 240
tacaatcata aagaaatcca acttaccttg tcacttattg acatgtgtgt gcagaactgt 300
ggtccaagtt tccagtctct gattgtgaag aaggaatttg ttaaagagaa tttagttaag 360
ctactgaatc ccagatacaa cttgccatta gacattcaga atagaatctt gaatttcatt 420
aagacttggt cacagggctt cccaggaggt gtggatgtaa gcgaagtcaa agaagtatac 480
ctcgacctgg ttaagaaagg cgttcagttt cctccctcag aagcagaggc tgaaacagca 540
agacaagaga ctgctcaaat ctcatcaaat cctccaacat ctgtccctac tgcaccagct 600
ctttcttctg taattgctcc aaagaactcg actgttacat tggtcccaga acagattgga 660
aaactgcaca gtgaattgga tatggtgaaa atgaatgtgc gagtgatgtc cgccatattg 720
atggagaata ctcctgggtc tgaaaaccat gaagacatag agcttctgca gaaactctat 780
aaaacaggtc gggagatgca ggagaggatc atggacctgc ttgtggtggt ggagaacgaa 840
gatgtaactg ttgagctaat tcaggtgaat gaggatttga ataatgctat ccttggatat 900
gagaggttta ctagaaacca acaaaggatt ttggagcaaa ataagaacca gaaggaagcc 960
accaatacta ccagtgagcc ttctgcccca tctcaagatc tcctcgacct aagtcccagt 1020
ccccggatgc ctagggccac tctgggagaa ctcaacacca tgaataatca actttcaggc 1080
ttaaatttca gccttccaag ttctgatgta acaaacaact taaaacccag tcttcatcca 1140
cagatgaact tgctagcctt ggagaataca gagataccc cgtttgccca aaggaccagc 1200
caaaacctca cctcaagcca cgcatatgat aattttctgg aacattcaaa ttcagtgttt 1260
ctacagccag ttagtctaca aaccattgca gcagcaccat caaaccagag tctgccacct 1320
ttgcccagca atcatccagc gatgacaaaa agtgatctcc agccacctaa ttactacgag 1380
gtaatggagt ttgatccctt agctcctgct gtcactacag aagctattta tgaagaaatt 1440
gatgctcacc agcacaaagg agctcaaaat gatggtgact gagaagaaag tggatgatca 1500
gctcactacc acatcaaagg tgccaactct ctaaaacgta gactctgtgc agctttgaag 1560
cctggaagac aatacctacc aacatgtcaa agccatggtg gcacatttct gctataatga 1620
agattaaata gaataacagt tccaggataa cactgattcc tgacaacagc atgagatttc 1680
aacagaactt gtttggaaca aatactcact taaaacttca gcagaagaaa aattacttag 1740
tccttaggcc aaccaattta actgcagtgt catgtttcac aggccttcct acatttagaa 1800
atcgtcacac agctgtgata agagtagatt attttactat gaaataattc tgaatagatg 1860
aaagcataaa atgtgagaaa ctgaatgtat tattcaggaa gaatactgag tgccttcatt 1920
taactaaagt tgaatgtaaa agtcaatttg cacttcttta taatcctctg gtttagaatt 1980
atgaaaaata gtggcaggct cgta 2004

<210> 36
<211> 4019
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2503986CB1

<400> 36
ccctggcgac agtgtcttgc cggggagtag tagccgggct ggtaactgga gtttgagatt 60
aggagacttt cagacccttg tgcacaaaga ggtgcggtcc ccgaggcagg atgaagttaa 120
aggaagtaga tcgtacagcc atgcaggcat ggagccctgc ccagaatcac cccatttacc 180
tagcaacagg aacatctgct cagcaattgg atgcaacatt tagtacaaat gcttcccttg 240
agatatttga attagacctc tctgatccat ccttggatat gaaatcttgt gccacattct 300
cctcttctca caggtaccac aagttgattt gggggcctta taaaatggat tccaaaggag 360
atgtctctgg agttctgatt gcaggtggtg aaaatggaaa tattattctc tatgatcctt 420
ctaaaattat agctggagac aaggaagttg tgattgccca gaatgacaag catactggcc 480
cagtgagagc cttggatgtg aacattttcc agactaatct ggtagcttct ggtgctaatg 540
aatctgaaat ctacatatgg gatctaaata attttgcaac cccaatgaca ccaggagcca 600
aaacacagcc gccagaagat atcagctgca ttgcatggaa cagacaagtt cagcatattt 660
tagcatcagc cagtcccagt ggccgggcca ctgtatggga tcttaggaaa aatgagccaa 720
tcatcaaagt cagtgaccat agtaacagaa tgcattgttc tgggttggca tggcatcctg 780
atgttgctac tcagatggtc cttgcctccg aggatgaccg gttaccagtg atccagatgt 840
gggatcttcg atttgcttcc tctccacttc gtgtcctgga aaaccatgcc agggggattt 900
tggcaattgc ttggagcatg gcagatcctg aattgttact gagctgtgga aaagatgcta 960
agattctctg ctccaatcca aacacaggag aggtgttata tgaacttccc accaacacac 1020
agtggtgctt cgatattcag tggtgtcccc gaaatcctgc tgtcttatca gctgcttcgt 1080

ttgatgggcg tatcagtgtt tattctatca tgggaggaag cacagatggt ttaagacaga 1140
aacaagttga caagctttca tcatcttttg ggaatcttga tccctttggc acaggacagc 1200
cccttcctcc gttacaaatt ccacagcaga ctgctcagca tagtatagtg ctgcctctga 1260
agaagccgcc caagtggatt cgaaggcctg ttggtgcttc tttttcattt ggaggcaaac 1320
tggttacgtt tgagaatgtc agaatgcctt ctcatcaggg agctgagcag cagcagcagc 1380
agcaccatgt gttcattagt caggttgtaa cagaaaagga gttcctcagc cgatcagacc 1440
aacttcagca ggctgtgcag tcacaaggat ttatcaatta ttgccaaaaa aaaattgatg 1500
cttctcagac tgaatttgag aaaaatgtgt ggtccttttt gaaggtaaac tttgaggatg 1560
attctcgtgg aaaatacctt gaacttctag gatacagaaa agaagatcta ggaaagaaga 1620
ttgctttggc cttgaacaaa gtggatggag ccaatgtggc tcttaaagac tctgaccaag 1680
tagcacagag tgatggggag gagagccctg ctgctgaaga gcagctcttg ggagagcaca 1740
ttaaagagga aaaagaagaa tctgaatttc taccctcatc tggaggaaca tttaatatct 1800
ctgtcagtgg ggacattgat ggtttaatta ctcaggcttt gctgacgggc aattttgaga 1860
gtgctgttga cctttgttta catgataacc gcatggccga tgccattata ttggccatag 1920
caggtggaca agaactcttg gctcgaaccc agaaaaaata cttcgcaaaa tcccaaagca 1980
aaattaccag gctcatcact gcagtggtga tgaagaactg gaaagagatt gttgagtctt 2040
gtgatcttaa aaattggaga gaggctttag ctgcagtatt gacttatgca aagccggatg 2100
aattttcagc cctttgtgat ctttgggaa ccaggcttga aaatgaagga gatagcctcc 2160
tgcagactca agcatgtctc tgctatattt gtgcagggaa tgtagagaaa ttagttgcat 2220
gttggactaa agctcaagat ggaagccacc ctttgtcact tcaggatctg attgagaaag 2280
ttgtcatcct gcgaaaagct gtgcaactca ctcaagccat ggacactagt actgtaggag 2340
ttctcttggc tgcgaagatg agtcagtatg ccaatttgtt ggcagctcag ggcagtattg 2400
ctgcagcctt ggcttttctt cctgacaaca ccaaccagcc aaatatcatg cagcttcgtg 2460
acagactttg tagagcacaa ggagagcctg tagcaggaca tgaatcacct aaaattccgt 2520
acgagaaaca gcagctcccc aagggcaggc ctggaccagt tgctggccac caccagatgc 2580
caagagttca aactcaacaa tattatcccc atggagaaaa tcctccacct ccgggtttca 2640
taatgcatgg aaatgttaat ccaaatgctg ctggtcagct tcccacatct ccaggtcata 2700
tgcacaccca ggtaccacct tatccacagc cacagcctta tcaaccagcc cagccgtatc 2760
ccttcggaac agggggtca gcaatgtatc gacctcagca gcctgttgct cctcctactt 2820
caaacgctta ccctaacacc ccttacatat cttctgcttc ttcctatact gggcagtctc 2880
agctgtacgc agcacagcac caggcctctt cacctacctc cagccctgct acttctttcc 2940
ctcctccccc ttcctctgga gcatccttcc agcatggcgg accaggagct ccaccatcat 3000
cttcagctta tgcactgcct cctggaacaa caggtacact gcctgctgcc agtgagctgc 3060
ctgcgtccca aagaacaggt cctcagaatg gttggaatga ccctccagct ttgaacagag 3120
tacccaaaaa gaagaagatg cctgaaaact tcatgcctcc tgttcccatc acatcaccaa 3180
tcatgaaccc gttgggtgac ccccagtcac aaatgctgca gcaacagcct tcagctccag 3240
taccactgtc aagccagtct tcattcccac agccacatct tccaggtggc cagcccttcc 3300
atggcgtaca gcaacctctt ggtcaaacag gcatgccacc atcttttca aagcccaata 3360
ttgaaggtgc cccaggggct cctattggaa ataccttcca gcatgtgcag tctttgccaa 3420
caaaaaaaat taccaagaaa cctattccag atgagcacct cattctaaag accacatttg 3480
aggatcttat tcagcgctgc ctttcttcag caacagaccc tcaaaccaag aggaagctag 3540
atgatgccag caaacgtttg gagtttctgt atgataaact tagggaacag acactttcac 3600
caacaatcac cagtggttta cacaacattg caaggagcat tgaaactcga aactactcag 3660
aaggattgac catgcatacc cacatagtta gcaccagcaa cttcagtgag acctctgctt 3720
tcatgccagt tctcaaagtt gttctcaccc aggccaataa gctgggtgtc taaaaggaca 3780
gcttctcttc cactcaatat tgccattttt ccaaagaaac atgttaaaaa aaaaaattat 3840
aagacatgga ctagtcctca ttagcatgtt tgcatagcaa ccagtcaaga gcatttacac 3900
tatttctgct gatatactca ccttagaact gctcagaacc ctggtgcttt atttttggtt 3960
ttaatctttt gttgccagtg atgattttcc tattctgcaa atagtgccac tccgcgact 4019

<210> 37
<211> 1793
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2596566CB1

<400> 37
cgacgccgga cgtgcggcag ttgcaggcga gcaggcgagg aatcgccgtg gcgtcttggt 60
gttctccacg ctggttcgca ggtgaagaga tggcgtttgt gaagagtggc tggttgctgc 120
gacagagtac tattttgaag cgctggaaga agaactggtt tgatctgtgg tcggatggtc 180
acctgatcta ttatgatgac cagactcggc agaatatcga ggataaggtc cacatgccaa 240
tggactgcat caacatccgc acggggcagg aatgtcggga tactcagccc ccggatggaa 300
agtcaaaaga ctgcatgctc cagattgttt gtcgagatgg gaaaacaatt agtctttgtg 360
cagaaagcac agatgattgc ttggcctgga aatttacact ccaagattct aggacaaaca 420

cagcgtatgt gggctctgca gtcatgaccg atgagacatc cgtggtttcc tcacctccac 480
catacacggc ctatgctgca ccggcccctg agcaggctta tggctatggg ccatacggtg 540
gtgcgtaccc gccaggaact caagttgtct acgctgcgaa tgggcaggcg tatgccgtgc 600
cctaccagta cccatatgca ggactttatg gacagcagcc tgctaaccaa gtcatcattc 660
gagagcgcta tcgagacaac gacagcgacc tggcactggg catgctggca ggagcagcca 720
cgggcatggc cttagggtct ctattttggg tcttctaggg gcctcaaggt cttgatgtgc 780
atagcttctg ataaccctgt gtgcaataat atgatttgca gggcatttct gtttgtgaca 840
aaagttttta ataatagttt taatcattcc tttgaaagta gtgatgtcat aattgtacta 900
atccacataa gtaccacaga gaagggtttg aactgtgcta ttttgttcaa atgttgactc 960
tccgggggca ctggctcatt ccaagactgt tcttgtgcaa ctctcagaat accttatttg 1020
agcatacctg ttttgaaagg cattttcttt ttagagttag gtgtagtgct taagggttaa 1080
tttattttca tgttatgcca gtaatatagt gttgtatgcc tattgagtga ttgtggcaag 1140
aaaagctaca gcttctttgc gtttaacttt ttcaaaccac agaccagaac tggttgcatg 1200
ttactttagg agttgtgggt tggtaagctc ccaggtactt cccgaggcta tggtgtgaga 1260
gccccgtcc tgccctctgg ggctccacag gcccctggca aggccgatgg ctcaggatga 1320
tggggcacag cccgcctttg aacaatcatg cttcagaaat ctgcctgacc ctagctgctg 1380
ctgctgctca ctttattctt gtatggcttt ggtaggcata cttggagaac atatcccaca 1440
ttaggaattg atttaagcct gagagtttga gggctttaat cctttaaaac ttggagaagc 1500
tggctgggcg cggtggctca cgcctgtaat cccagcactt tgagagaccg aggcgggcgg 1560
atcacgaggt caggagatcg agaccatcct ggctaacacg gtgaaacccc atctctacta 1620
aaaatacaaa aaattagctg ggcgtggtgg caggcgcctg tggtcccagc tactcgggag 1680
gctgaggcag gagaatagtg tgaacccagg aggcggagct tgcagtgagc caagatagtg 1740
ccactgcact tcagcctggg tgacagagtg agactctgtc tcaaaaaaaa aaa 1793

<210> 38
<211> 4427
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2685253CB1

<400> 38
cttcggcgag acctaccggg ccctgtctgt gccggccgtg agctgcggca ggacggctgg 60
agattacttc tctaacagga tcaccagttt gctccacaaa gcacaatgtc tcgatcacga 120
caaccccccc ttgtgaccgg catctctcca aatgaaggga taccatggac gaaggtcaca 180
atcaggggag aaaatctggg gactggcccc accgacctca taggcttgac catttgtgga 240
cataattgcc tcctgacggc agaatggatg tctgcaagta aaatagtatg tcgagtggga 300
caagccaaaa atgacaaagg agacattatt gtcaccacta agtcaggtgg cagaggaacc 360
tcaacagtct ctttcaagct actcaaacct gagaaaatag gcattttgga tcagtctgct 420
gtgtgggttg atgaaatgaa ttattatgat atgcgcactg acaggaacaa aggaattccg 480
cccttgtcct tacgtcctgc taacccgctt ggcattgaga ttgaaaaaag taaattttcg 540
cagaaggact tagaaatgct attccatgga atgagtgctg attttacaag tgagaatttc 600
tcagcagcct ggtatcttat agagaatcac tcaaacacca gttttgagca gctcaaaatg 660
gcagtcacca acctaaagag acaggctaac aagaagagtg agggcagcct ggcctatgtg 720
aaaggcggtc tcagtacatt cttcgaagca caggatgccc tctcagccat ccatcaaaaa 780
ctagaagcag atggaacgga aaaagtagaa ggatccatga cgcagaaact ggagaatgtt 840
ctgaacagag caagtaatac tgcagacaca ttgtttcaag aagtattagg tcggaaagac 900
aaggcagatt ccactagaaa tgcactcaat gtgcttcagc gatttaagtt tcttttcaac 960
cttcctctaa atattgaaag gaatattcaa aagggtgatt atgatgtggt tattaatgat 1020
tatgaaaagg ccaagtcact ttttgggaaa acggaggtgc aagttttcaa gaaatattat 1080
gctgaagtag aaacaaggat tgaagcttta agagaattac ttctggataa attgcttgag 1140
acaccatcaa ctttacatga ccaaaaacgt tacataaggt acctgtctga ccttcatgcg 1200
tctggtgacc ctgcttggca atgcattgga gcccaacaca agtggatcct tcagctcatg 1260
cacagttgca aagagggcta cgtgaaagat ctgaaaggta acccaggcct gcacagtccc 1320
atgttggatc ttgataatga tacacgtccc tcagtgttgg gccatctcag tcagacagcg 1380
tccctgaaga ggggcagcag ctttcagtct ggtcgagacg acacgtggag atacaaaact 1440
ccccacaggg tggcctttgt tgaaaaattg acaaaactcg tcttgagcca gctgcctaac 1500
ttctggaaac tctggatctc ctacgttaat ggaagcctct tcagtgagac tgctgagaag 1560
tcaggccaga ttgaaagatc aaagaatgta aggcaaagac aaaatgattt taagaaaatg 1620
attcaggaag taatgcactc cctggtgaag cttacccgcg gagccctgct tcccctcagc 1680
atccgggatg gggaagccaa gcagtacgga ggctgggagg tgaagtgcga gctctccgga 1740
cagtggctcg ctcacgccat ccagactgta agacttactc atgaatcgtt gactgccctt 1800
gaaattccta atgacctgtt acagactatc caggatctca tcttggatct ccgagtacgt 1860
tgcgtaatgg ccacgttgca gcacacggcg gaagaaataa agagattagc tgaaaaagaa 1920
gactggattg ttgacaatga aggactgact tctctaccat gtcagtttga acagtgcatc 1980

gtgtgttctc tgcagtcact gaaggggttt ctggagtgca agccgggaga ggccagtgtc 2040
ttccaacaac ctaaaacaca ggaggaggtt tgccagctaa gcatcaatat aatgcaggtt 2100
tttatatact gtctggaaca gttgagcacc aagcctgatg cagatataga tactacacat 2160
ctctctgttg atgtttcttc ccctgacttg tttggaagta tccatgaaga cttcagcttg 2220
acctcagaac agcgcctttt gatagtccta agtaattgct gctatctaga acgtcacacc 2280
ttcctaaata tcgcagaaca ttttgaaaag cacaacttcc agggaataga aaaaatcaca 2340
caggttagca tggcctcatt gaaagaacta gatcaaagac tctttgaaaa ttacatcgag 2400
ttgaaagcag atcccatcgt tggctcctta gaacctggaa tttatgcagg atattttgat 2460
tggaaggact gcctgcctcc aacaggtgtc agaaactatt taaaagaagc actggtgaat 2520
ataattgccg tgcatgcaga ggtgttcacc atttccaaag aactggtccc tcgggtacta 2580
tccaaggtga tagaagcagt ttctgaagag ctcagtcgac tgatgcagtg tgtttcatcc 2640
ttcagcaaaa atggagcttt acaggcgaga cttgaaatct gtgctttgag ggacactgtg 2700
gctgtttacc tgacacccga aagcaagtca agttttaagc aggctttgga agccctgccc 2760
cagctttcca gtggagcaga taaaaagtta ctggaagagc tcctgaacaa gttcaagagt 2820
agcatgcact tgcagctcac ctgtttccaa gcagcttctt caaccatgat gaaaacataa 2880
atatctgcca cataaaagaa gtccaggaaa ataacacgta ataagactgt tcactctcta 2940
agtaccctaa aggtatttgg tgtattaaac attgggtttg ccatttttct ctttttctt 3000
cctctgactt cgaaaattgt tggtatacat ttcaaccaaa atgacctcat ttgaaatgcc 3060
caggaagtat ttgttttgcc attcttacta gatcagatcc tgtatgactt taaaatacat 3120
tttaaaatat attttgcatc tcagcagaag gttgagtgct gcaggaagag ctgcttctgc 3180
agggagtttt ctggaatggc gtggcacata ttgaagcatt tctcagtgtc ccgttcacgg 3240
aagcgcaggc agctcccacg acacacggag agactgactg atacgagatt tggaaagcta 3300
tgtagacatc tttggagctc ttactgtcct aaactgtaca gctgtgctta aaacccttat 3360
ttcatataaa tggccttaag ttttctaatt caagcgggtt tttggaaaaa tttatggtct 3420
ccattaaaat acatattaca actggggtag attatttgtg gtccaggtgt ctgtgattta 3480
actttgcgtt ttgctatctg atttttattt ttcacagggc taagcatgag ctttcattct 3540
cactcactct taatttgtcg tgcgtcacta cacatgcacg tgttgcagtc cctgaggccc 3600
tgtgtgttat ctgtgatgga gtgtgaatgt gtaacgggca ctgtgttaca ctctcaggtg 3660
ttggcggggc ggtcgcagac ttcagggtcc cctaacggaa aggccaggct ccgcgtggac 3720
ggccaactcc ctgcccgctc cttcagcagg tgactgtctc tgccacttct tacctgctga 3780
aggatcttgc tcagtagctg gaacaatgct gctgtcacac agtctcttct ctgaaacttc 3840
aggatgctcc ttggtcacca ggcaatggga gctgtagacc agccgcatgc acttgcccca 3900
cattcactgc tgactggctt cactggaata ggttcaggtc accgggactt ctctcagagt 3960
cagcagccca cccaactccc tcataccgtc gcatctgaaa attttcagag aggaattctc 4020
tttgtagtcg gcttgtcagg gttttcaagt ttttctatgg ttatttttaa atctcttttt 4080
ttaaatgcta aaattaatgt cctcattaat gcaagtaatt ttagagcaca gtgtagccat 4140
gtacagtttc cattatttaa aataaataaa tgtgttggtt catttggtgt atgggggggac 4200
atcatggtaa ctggaaatga aagtttaatc tttcacaaat gtgcttgata aagctgttaa 4260
gaattgtctt tcctaagtat gttaactgtt gtatctaact caactgtgag ctaagtttaa 4320
aaactgcaat catgaatgtt tgtgatattt taaatatcta ctacagcaca ttcagccaat 4380
aaagaatctg cctttaatgt aagttgtaat ttacttgaaa aaaaaaa 4427

<210> 39
<211> 1583
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2762252CB1

<400> 39
gagctgaaag ccaggaacat ccgaggagaa gagaaagctt ccagccctcc tcccttcacc 60
ctggaaatcc agacaccccc acccccaccc tcagatcact ttaagataat ttctttattc 120
gtttgcccga cagaccatgg ctccctttgg aagaaacttg ctaaagactc ggcataaaaa 180
cagatctcca actaaagaca tggattcaga agagaaggaa attgtggttt gggtttgcca 240
agaagagaag cttgtctgtg ggctgactaa acgcaccacc tctgctgatg tcatccaggc 300
tttgcttgag gaacatgagg ctacgtttgg agagaaacga tttcttctgg ggaagcccag 360
tgattactgc atcatagaga agtggagagg ctccgagagg gttcttcctc cactaactag 420
aatcctgaag ctttggaaag cgtggggaga tgagcagccc aatatgcaat ttgttttggt 480
taaagcagat gcttttcttc cagttccttt gtggcggaca gctgaagcca aattagtgca 540
aaacacagaa aaattgtggg agctcagccc agcaaactac atgaagactt taccaccaga 600
taaacaaaaa agaatagtca ggaaaacttt ccggaaactg gctaaaatta agcaggacac 660
agtttctcat gatcgagata atatggagac attagttcat ctgatcattt cccaggacca 720
tactattcat cagcaagtca agagaatgaa agagctggat ctggaaattg aaaagtgtga 780
agctaagttc catcttgatc gagtagaaaa tgatggagaa aactatgttc aggatgcata 840
tttaatgccc agtttcagtg aagttgagca aaatctagac ttgcagtatg aggaaaacca 900

gactctggag gacctgagcg aaagtgatgg aattgaacag ctggaagaac gactgaaata 960
ttaccgaata ctcattgata agctctctgc tgaaatagaa aaagaggtaa aaagtgtttg 1020
cattgatata aatgaagatg cggaagggga agctgcaagt gaactggaaa gctctaattt 1080
agagagtgtt aagtgtgatt tggagaaaag catgaaagct ggtttgaaaa ttcactctca 1140
tttgagtggc atccagaaag agattaaata cagtgactca ttgcttcaga tgaaagcaaa 1200
agaatatgaa ctcctggcca aggaattcaa ttcacttcac attagcaaca aagatgggtg 1260
ccagttaaag gaaaacagag cgaaggaatc tgaggttccc agtagcaatg gggagattcc 1320
tccctttact caaagagtat ttagcaatta cacaaatgac acagactcgg acactggtat 1380
cagttctaac cacagtcagg actccgaaac aacagtagga gatgtggtgc tgttgtcaac 1440
atagttccaa tggctccttt ctgacctgct ttcatgtttt aatgtttgtt taatttaata 1500
ggaaacctca ttttaaatat aacactcaaa aaaaatgtaa atcatattgt agtattcaat 1560
agttaataaa aactcgagaa atg 1583

<210> 40
<211> 1416
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3452009CB1

<400> 40
gtccggcggg ttacagcgga ggcctaggtg gcagacaggg ggcccgggcc gctgcgtgtt 60
gtccacccaa gatggagttc ctcctgggga acccgttcag cacaccagtg gggcagtgcc 120
tcgaaaaggc aacagatggc tccctgcaaa gtgaggattg gacgttgaat atggagatct 180
gtgacatcat caatgagacg gaggaagggc caaaggatgc cattcgagcc ctgaagaagc 240
ggctcaacgg gaaccggaac tacagagagg tgatgctggc attaacagtg ctggagacat 300
gtgtgaagaa ctgtggccac cgcttccaca tccttgtggc caaccgagat ttcatcgaca 360
gtgttctggt caaaattata tctcccaaga acaaccctcc caccattgta caggacaaag 420
tgcttgctct gatccaggca tgggctgatg cctttcgaag cagtcctgat ctcaccggcg 480
ttgtgcacat atatgaggag ctgaagagga aaggggttga atttcccatg gcagacttgg 540
acgctctgtc tcccatacac acaccacagc ggagtgtccc tgaagtggat ccagctgcga 600
ccatgcccag gtcccaatca cagcagagga caagtgctgg ttcctattcc tcgccgcctc 660
ctgctcccta ctccgcaccg caggccccag ctctgagtgt gactggcccc atcacagcca 720
attcagaaca gattgccagg ctgcggagtg aactggacgt cgttcgagga aacacaaaag 780
tcatgtctga gatgttaaca gaaatggtcc ctggacagga ggattcatct gatctggagt 840
tgctgcagga gctcaacagg acctgtcggg ccatgcagca gcgcatcgtg gagctcatct 900
cccgcgtgtc caatgaggag gtcaccgagg agctgctgca tgtgaacgat gacctcaaca 960
acgtcttcct tcgatacgag aggtgggagc cggatttctt ttttttttt tttcccttga 1020
aaagattgtt gccataacag agggatttct ttcataaaca aaggaaagag agggagaggg 1080
agagagatta gagttggaaa atactcatca gagatgctaa tttgttcccg gctgacagga 1140
aaaatgcttc ttgtcctcct gggaactgag cagcagaagc agcaggaagt tgacgtgcag 1200
aattaggctc tgccctttgt aagcctcagg ctgctctgcc tttccccact gccccgccac 1260
ccccagcccc agggaggggt agcacttggg ggcctctggg ccccagagct tctgcacctg 1320
ctcctttgat gatcctttga atttgaaaag gattaaaggg aaaaaaaatc tctagtaaga 1380
aaaaggattt cctctttccc atctgtgaaa aaaaaa 1416

<210> 41
<211> 1662
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4644780CB1

<400> 41
cgatgccggc ggtcagtggt ccaggtccct tattctgcct tctcctcctg ctcctggacc 60
cccacagccc tgagacgggg tgtcctcctc tacgcaggtt tgagtacaag ctcagcttca 120
aaggcccaag gctggcattg cctggggctg gaataccctt ctggagccat catggagacg 180
ccatcctggg cctggaggaa gtgcggctga cgccatccat gaggaaccgg agtggcgccg 240
tgtggagcag ggcctctgtc cccttctctg cctgggaagt agaggtgcag atgagggtga 300
cgggactggg gcgccgggga gcccagggca tggccgtgtg gtacacccgg ggcaggggcc 360
atgtaggctc tgtccttggg gggctggctt cgtgggacgg catcgggatc ttctttgact 420
ctccggcaga ggatactcag gacagtcctg ccatccgtgt gctggccagc gacgggcaca 480
tcccctctga gcagcctggg gatggagcta gccaagggct gggctcctgt cattgggact 540

tccggaaccg gccacacccc ttcagagcac ggatcaccta ctgggggcag aggctgcgca 600
tgtccttgaa cagtggcctc actcccagtg atccagatga tcatgatgtc ctgtccttcc 660
tgaccttcag cctgagtgag cccagcccag aggttccccc tcagcccttc ctggagatgc 720
agcagctccg cctggcgagg cagctggaag ggctgtgggc aaggctgggc ttgggcacca 780
gggaggatgt aactccaaaa tcagactctg aagctcaagg agaaggggaa aggctctttg 840
acctggagga gacgctgggc agacaccgcc ggatcctgca ggctctgcgg ggtctctcca 900
agcagctggc ccaggctgag agacaatgga agaagcagct ggggcccccа ggccaagcca 960
ggcctgacgg aggctgggcc ctggatgctt cctgccagat tccatccacc ccagggaggg 1020
gtggccacct ctccatgtca ctcaataagg actctgccaa ggtcggtgcc ctgctccatg 1080
gacagtggac tctgctccag gccctgcaag agatgaggga tgcagctgtc cgcatggctg 1140
cagaagccca ggtctcctac ctgcctgtgg gcattgagca tcatttctta gagctggacc 1200
acatcctggg cctcctgcag gaggagcttc ggggcccggc gaaggcagca gccaaggccc 1260
cccgcccacc tggccagccc ccaagggcct cctcgtgcct gcagcctggc atcttcctgt 1320
tctacctcct cattcagact gtaggcttct tcggctacgt gcacttcagc aggcaggagc 1380
tgaacaagag ccttcaggag tgtctgtcca caggcagcct tcctctgggt cctgcaccac 1440
acacccccag ggccctgggg attctgagga ggcagcctct ccctgccagc atgcctgcct 1500
gacccacctc agagcctgct ttgcatcact gggaagcagg cagtgtcttg ggtgggggct 1560
tggtcagtat cctctccgtc tgggtgccca gctcccacgc acacctgagc tttcggcatg 1620
ctcccacctc gttaaaggtg atttccctct caaaaaaaaa aa 1662

<210> 42
<211> 3387
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4946103CB1

<400> 42
ccggtctccg ttttggaaga cccgcctcgg cacagccagg ctcagtccgg ccttgcggta 60
agccttcggc cgcggctgcc cggtagtccc ggcggcggcg gacagacgag ctgacaggca 120
ccagggtcta aggcggctcc tcagtccggc tgctgtctcc acgcctgggg tcgggcaccg 180
ctccttctga ccttcctttc cccgtttgtc ccgctgagaa aagatgacag caatcaagca 240
tgcattacaa agagacattt ttacaccaaa tgatgaacgc ctgctgagca ttgtgaatgt 300
ctgcaaagca ggaaaaaaga aaaagaactg ttttttatgt gccacagtga caactgaacg 360
ccctgtgcag gttaaggtgg tcaaagtcaa gaaatccgat aagggagatt tctacaaaag 420
gcagattgca tgggcccttc gagatcttgc tgtggtagat gccaaagatg ctatcaaaga 480
aaatcctgaa tttgatttac actttgaaaa aatatataaa tgggttgcca gcagcactgc 540
tgaaaagaat gcatttattt catgcatttg gaaattgaat cagcgatatc tccggaagaa 600
aattgatttt gtcaatgtta gctcacagct tttggaagaa tctgttccaa gtggagaaaa 660
tcagagtgtg acaggaggtg atgaagaagt agtagatgaa taccaagagt taaatgcaag 720
agaagaacag gatatcgaaa taatgatgga aggctgtgaa tatgcaatct cgaatgcgga 780
acgctttgca gaaaaattgt ccagagagct gcaggtgcta gatggggcta acatccagtc 840
aatcatggca tctgaaaaac aagtcaacat cctgatgaaa ttgctagatg aggctctaaa 900
ggaggtagat cagattgaat tgaaactgag cagttatgag gaaatgctcc aaagtgtaaa 960
agaacaaatg gatcagatct ctgaaagcaa ccacctaatt catcttagta acactaataa 1020
tgtaaaactc ctatctgaga tagagttcct tgtgaaccac atggacttgg ccaaaggtca 1080
tataaaggcc cttcaggaag gagatcttgc ttcttccaga ggcattgagg cctgcaccaa 1140
tgctgctgat gcccttctgc agtgcatgaa tgtagctctt cgaccaggcc atgacttgct 1200
tctggcagtc aaacagcaac agcagcgatt cagtgatttg cgagagcttt ttgcccggag 1260
actggccagt cacctcaaca atgtttttgt tcaacagggc cacgatcaga gttcgtctct 1320
cccccagcac tgtgtctcaa ccgggtttac ccaatcatca tccatttcac agagatttcc 1380
tccgattgcc aagctgatgg agtggctaaa gagtacagat tatggaaaat atgaaggact 1440
aacaaagaat tacatggatt atttatcccg actatatgaa agagaaatca aagatttctt 1500
tgaagttgca aagatcaaga tgactggcac aactaaagaa agcaagaagt ttggtcttca 1560
tggaagttcg gggaaattaa ctggatctac ttctagtcta aataagctca gtgttcagag 1620
ttcagggaat cgcagatctc agtcatcttc cctgttggat atgggaaaca tgtctgcctc 1680
tgatctcgat gttgccgaca ggaccaaatt tgataagatc tttgaacagg tactaagtga 1740
actggagccc ctatgtctgg cagaacagga cttcataagt aaatttttca aactacagca 1800
acatcaaagt atgcctggaa ctatggctga agcagaggac ctggatggag gaacattatc 1860
acggcaacat aattgtggca caccactgcc tgtttcatct gagaaagata tgatccgcca 1920
aatgatgatt aaaatatttc gctgcattga gccagagctg aacaacctaa ttgcattagg 1980
agacaaaatt gatagcttta actctcttta tatgttagtc aaaatgagtc atcatgtgtg 2040
gactgcacaa aatgtggacc ctgcttcttt cctaagtact acattgggaa atgttttggt 2100
gactgtcaaa aggaactttg acaaatgcat tagtaaccaa ataaggcaaa tggaagaagt 2160
aaagatctca aaaaagagta aagttggaat tcttccattt gttgctgaat ttgaagaatt 2220

tgctggactt gcagaatcaa tcttcaaaaa tgctgagcgt cgtggagacc tggataaagc 2280
atacaccaaa cttatcagag gagtatttgt taatgtggag aaagtagcaa atgaaagcca 2340
gaagaccccc agggatgtgg ttatgatgga aaactttcac catatttttg caactctttc 2400
tcgattgaaa atctcatgtc tagaagcaga aaaaaaagaa gccaaacaaa aatacacaga 2460
tcaccttcag tcttatgtca tttactcttt aggacaacct cttgaaaaac taaatcattt 2520
ctttgaaggt gttgaagctc gcgtggcaca gggcataagg gaggaggaag taagttacca 2580
acttgcattt aacaaacaag aacttcgtaa agtcattaag gagtaccctg gaaaggaagt 2640
aaaaaaaggt ctagataacc tctacaagaa agttgataaa catttatgtg aagaagagaa 2700
cttacttcag gtggtgtggc actccatgca agatgaattt atacgccagt ataagcactt 2760
tgaaggtttg atagctcgct gttatcctgg atctggtgtt acaatggaat tcactattca 2820
ggacattctg gattattgtt ccagcattgc acagtcccac taaaccttgt gaaagaagaa 2880
aagataactg aatgaagcat ttgagtataa cagacactat accaaaatac caagcaactg 2940
ttttgagaac ccagacttaa aattttatgt attattaaat gttagataaa tgggtagtac 3000
catactacaa atatttaaat gcaaaattac caacctatat agcagtttta tttgccctat 3060
aggttgcata ctaacttaag cattcatgtc accataaaat gcctttagca tttctcaatg 3120
actggatggg aaatttтcct ttattgccta gctgcttgtg tttgagtggt tgtcctatga 3180
gcaatgcatt tggagttctt cagctttcac tacttctctg ttgcttgcta atcatgtaac 3240
tactaaaata ctgtacaaaa ttgtttttтc acactaacaa atgtgtatat ggagaagagg 3300
gctcatgtga tgatcatttg tgaacttaga tttttgagga ttatgtgact agtaataaat 3360
gtgaaataaa ttttcaaaaa aaaaaaa 3387

<210> 43
<211> 1043
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5562355CB1

<400> 43
gcgggcgaca tggacaacgc ggggaaggag cgtgaggcag tacagctgat ggcggaggcc 60
gagaagcgag tcaaggcctc ccactccttc ctccgagggc tgtttggagg aaacacaaga 120
atagaagagg cttgtgaaat gtataccaga gctgcaaata tgttcaagat ggctaaaaat 180
tggagtgctg caggaaacgc attttgtcag gcagccaagc tccacatgca gcttcagagc 240
aaacatgact ctgctaccag ctttgtggat gctggaaatg cttacaaaaa ggcagatccc 300
caagaggcta tcaactgctt aaatgcagcc atcgacattt acacagacat gggaaggttt 360
acaattgcag ccaagcacca cattactatt gcagagatct atgagactga acttgtagac 420
attgagaagg ctattgcaca ttatgaacaa tctgctgatt attacaaagg agaagaatcc 480
aacagctcag caaacaagtg tctgctgaag gtggcagcat atgctgccca tcttgagcag 540
taccagaacg ccattgagat ctatgagcag gttggggcaa acacaatgga taatcctttg 600
acgacataca gtgcaaagga ttacttcttc aaagctgccc tctgccactt catagtagac 660
gagttgaatg ccaagcttgc tcttgagcaa tatgaggaca tgtttccagc atttactgat 720
tcaagagaat gtaaattatt gaaaaaactc ctagaagctc atgaagaaca gaacagtgaa 780
gcttacactg aagcagtgaa ggaatttgac tcaatatctc gcttggatca gtggctgacc 840
accatgttgc ttcgcatcaa aaagtccatc caaggggatg gagaaggaga tggagaccta 900
aaatgaaatg tttttgtctt tgtggcatgc agctaactcc tctttagttt tgtcttaggg 960
tcaagtgatc tttatgggat gcctatttaa tggcttaatt ttgttgcata tgagccagac 1020
ggcctgtgta ttgtttaagc tcg 1043

<210> 44
<211> 3207
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5678824CB1

<400> 44
ccggaggtgg gagccctggg ccaaaatggc ggcctacctg cagtggcggc gcttcgtttt 60
cttcgacaag gagctggtga aggagccgct gagcaatgat ggggccgctc ccggggccac 120
acctgcttct ggatccgctg cttccaagtt cctttgcctc cctcctggca tcactgtctg 180
cgactcaggc cgagggagcc tggtctttgg agatatggaa ggccagatct ggttcttgcc 240
acgttcccta cagcttacag gcttccaagc ctacaaacta cgggtgacac acctgtacca 300
actgaagcag cacaatattc tggcatctgt tggagaagat gaagagggca tcaacccctt 360
ggttaagatc tggaacctgg agaagagaga tggtggcaat ccactctgca ctcgaatctt 420

ccctgctatt ccaggaacag agccaactgt tgtatcttgt ttgactgtcc atgaaaatct 480
caactttatg gccattggtt tcacagatgg cagtgttaca ttgaacaaag gagacatcac 540
ccgggaccgg catagcaaga cccagatttt gcacaagggc aactatcctg taactggatt 600
ggcctttcgc caagcaggaa agaccactca cttgtttgtt gtgacaacag agaacgtcca 660
gtcctatata gtttctggaa aagactaccc tcgcgtggag ttggacaccc atggttgtgg 720
cctgcgctgc tcagccctaa gtgacccttc tcaggacctg cagttcattg tggccgggga 780
tgagtgtgtc tacttgtacc agcctgatga acgtgggccc tgcttcgcct ttgagggcca 840
taagctcatt gcccactggt ttagaggcta ccttatcatt gtctcccgtg accggaaggt 900
ttctcccaag tcagagttta ccagcaggga ttcacagagc tccgacaagc agattctaaa 960
catctatgac ctgtgcaaca agttcatagc ctatagcacc gtctttgagg atgtagtgga 1020
tgtgcttgct gagtggggct ccctgtacgt gctgacgcgg gatgggcggg tccacgcact 1080
gcaggagaag gacacacaga ccaaactgga gatgctgttt aagaagaacc tatttgagat 1140
ggcgattaac cttgccaaga gccagcatct ggacagtgat gggctggccc agattttcat 1200
gcagtatgga gaccatctct acagcaaggg caaccacgat ggggctgtcc agcaatatat 1260
ccgaaccatt ggaaagttgg agccatccta tgtgatccgc aagtttctgg atgcccagcg 1320
cattcacaac ctgactgcct acctgcagac cctgcaccga caatccctgg ccaatgccga 1380
ccataccacc ctgctcctca actgctatac caagctcaag gacagctcga agctggagga 1440
gttcatcaag aaaaagagtg agagtgaagt ccactttgat gtggagacag ccatcaaggt 1500
cctccggcag gctggctact actcccatgc cctgtatctg gcggagaacc atgcacatca 1560
tgagtggtac ctgaagatcc agctagaaga cattaagaat tatcaggaag cccttcgata 1620
catcggcaag ctgccttttg agcaggcaga gagcaacatg aagcgctacg gcaagatcct 1680
catgcaccac ataccagagc agacaactca gttgctgaag ggactttgta ctgattatcg 1740
gcccagcctc gaaggccgca gcgataggga ggccccaggc tgcagggcca actctgagga 1800
gttcatcccc atctttgcca ataacccgcg agagctgaaa gccttcctag agcacatgag 1860
tgaagtgcag ccagactcac cccaggggat ctacgacaca ctccttgagc tgcgactgca 1920
gaactgggcc cacgagaagg atccacaggt caaagagaag cttcacgcag aggccatttc 1980
cctgctgaag agtggtcgct tctgcgacgt ctttgacaag gccctggtcc tgtgccagat 2040
gcacgacttc caggatggtg tcctttacct ttatgagcag gggaagctgt tccagcagat 2100
catgcactac cacatgcagc acgagcagta ccggcaggtc atcagcgtgt gtgagcgcca 2160
tggggagcag gacccctcct tgtgggagca ggccctcagc tacttcgctc gcaaggagga 2220
ggactgcaag gagtatgtgg cagctgtcct caagcatatc gagaacaaga acctcatgcc 2280
acctcttcta gtggtgcaga ccctggccca caactccaca gccacactct ccgtcatcag 2340
ggactacctg gtccaaaaac tacagaaaca gagccagcag attgcacagg atgagctgcg 2400
ggtgcggcgg taccgagagg agaccacccg tatccgccag gagatccaag agctcaaggc 2460
cagtcctaag attttccaaa agaccaagtg cagcatctgt aacagtgcct tggagttgcc 2520
ctcagtccac ttcctgtgtg gccactcctt ccaccaacac tgctttgaga gttactcgga 2580
aagtgatgct gactgcccca cctgcctccc tgaaaaccgg aaggtcatgg atatgatccg 2640
ggcccaggaa cagaaacgag atctccatga tcaattccag catcagctca ggtgctccaa 2700
tgacagcttt tctgtgattg ctgactactt tggcagaggt gttttcaaca aattgactct 2760
gctgaccgac cctcccacag ccagactgac ctccagcctg gaggctgggc tgcaacgcga 2820
cctactcatg cactccagga ggggcactta agcagcctgg aggaagatgt gggcaacagt 2880
ggaggaccga gagaacagac acaatgggac ctgggcgggc gttacacaga aggctggctg 2940
acatgcccag ggctccactc tcatctaatg tcacagccct cagaactaaa gcggactttc 3000
tttccctgcc ttcttattta gtcagcttgc catccctcct cttcactagc agtgtagatc 3060
attccagatc agtgggggag ggcacctcag caacctctga gtgtggacaa tagctgcttt 3120
cttctctatc caagagcacc aggctgtgct tgggtccttg ctctcagagt ctataaataa 3180
aagaatataa tgattaaaaa aaaaaaa 3207

<210> 45
<211> 2356
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5870962CB1

<400> 45
ctgaactaac cctctaaccc ttgggagtct gtgtgcagca gtgatgtgag ctgcatccgg 60
gctgaaatgg gagctcccgc tatggcactg gctcaaggag ctaatctaaa gagaaagtat 120
tttggggaag acatcttggt tgcacccctc attgcatctt gactaattgc ttccaactcc 180
gggggcttcc agggacaatc tcattgttct tttctctcca agggtcagcc atgttaacca 240
tttgttaatt tcttctgaac agatcaagca aaagtagata atcagccaga agaattagtg 300
cgtagtgctg aagatgatga gaaaccagat cagaagccag ttacaaatga atgcgtacca 360
agaatttcca cagtgcctac acaacctgat aatccatttt ctcaccctga caaactcaaa 420
aggatgagca agtctgttcc agcatttctc caagatgagg tgagtggcag tgtgatgagt 480
gtttatagtg gagactttgg caatctggaa gttaaaggaa atattcagtt tgcaattgaa 540

tatgtggagt cactgaagga gttgcatgtt tttgtggccc agtgtaagga cttagcagca 600
gcggatgtaa aaaaacagcg ttcagaccca tatgtaaagg cctatttgct accagacaaa 660
ggcaaaatgg gcaagaagaa aacactcgta gtgaagaaaa ccttgaatcc tgtgtataac 720
gaaatactgc ggtataaaat tgaaaaacaa atcttaaaga cacagaaatt gaacctgtcc 780
atttggcatc gggatacatt taagcgcaat agtttcctag gggaggtgga acttgatttg 840
gaaacatggg actgggataa caaacagaat aaacaattga gatggtaccc tctgaagcgg 900
aagacagcac cagttgccct tgaagcagaa aacagaggtg aaatgaaact agctctccag 960
tatgtcccag agccagtccc tggtaaaaag cttcctacaa ctggagaagt gcacatctgg 1020
gtgaaggaat gccttgatct accactgcta aggggaagtc atctaaattc ttttgttaaa 1080
tgtaccatcc ttccagatac aagtaggaaa agtcgccaga agacaagagc tgtagggaaa 1140
accaccaacc ctatcttcaa ccacactatg gtgtatgatg ggttcaggcc tgaagatctg 1200
atggaagcct gtgtagagct tactgtctgg gaccattaca aattaaccaa ccaatttttg 1260
ggaggtcttc gtattggctt tggaacaggt aaaagttatg ggactgaagt ggactggatg 1320
gactctactt cagaggaagt tgctctctgg gagaagatgg taaactcccc caatacttgg 1380
attgaagcaa cactgcctct cagaatgctt ttgattgcca agatttccaa atgagcccaa 1440
attccactgg ctcctccact gaaaactact aaaccggtgg aatctgatct tgaaaatctg 1500
agtaggtgga caaatatcct cactttctat ctattgcacc taaggaatac tacacagcat 1560
gtaaaagtca atctgcatgt gcttctttga ttacaaggcc caagggattt aaatataaca 1620
aaatgtgtaa tttgtgactc taatattaaa taagatattt gaacaagcta ggaaaattga 1680
atttctgctg ctgcttcaaa gaaaaagctg ccccagagca ttaaacatgg ggtattgtta 1740
agaagcaaaa tgttcttgtt tgccatcatg tgtttcacac cacaattctg tgccacagtt 1800
aagagggtct ggtacccttg caggaccttt gtaggttgtg ggaaaaagtc gcagaaagat 1860
actcaaagtg gagcagggaa tggagacaga catcagtgat gataaaaaaa aaaaatggac 1920
cttaagaaac tatttactct gtaatctcta ataaaatatg gaattccata ttagggcaat 1980
gagactgaaa ctactggtgt ttttctgcct tgagaaaaca aacagttaaa acaagcctca 2040
aatgtatttt agtgccaccc actggccata ggtacaattc agttgttggc ttgttttgac 2100
ttaattctaa aataggtctc aagcctgtat ttttatgagt ttatttttt aaaaccctgc 2160
atatatatga ttgtttttct tataacttta ctatatgaaa gcagcataag agtagtcaca 2220
aacatgtttt gcaacaaagt tttaattaga atgtaagttg ctcagttata ctgttcttct 2280
tatgtatgta aaattttcgt attttgtaaa aacccttaga ataaattatc atttgattta 2340
aattgtatta gaaaat 2356

<210> 46
<211> 1034
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2818605CB1

<400> 46
gctttgctgt tgttgctctt cggaggcggc gatccccgaa ggcgagctga aatacggctg 60
caggctacaa tttgcagccg accattatgg atgacaagga gccgaagagg tggcccaccc 120
tcagggaccg cttgtgctcg gatggcttct tatttcccca ataccccatt aaaccgtatc 180
atctgaaggg gatccacaga gctgtcttct atcgtgatct ggaggaactg aagttcgttc 240
tgctcacgcg ttatgacatc aataagagag acaggaagga aaggaccgcc ctacatttgg 300
cctgtgccac tggccaaccg gaaatggtac atctcctggt gtccagaaga tgtgagctta 360
acctctgcga ccgtgaagac aggacacctc tgatcaaggc tgtacaactg aggcaggagg 420
cttgtgcaac tcttctgctg caaaatggcg ccgatccaaa tattacggat gtctttggaa 480
ggactgctct gcactacgct gtgtataatg aagatacatc catgatagaa aaacttcttt 540
cacatggtac aaatattgaa gaatgcagca aggtataggt caaccaatgt tattttcaaa 600
ctatctgaaa tgcatttatt ttaacattga cacatgtaag ggtcaatttt tcatatttgg 660
aagctcaaac attccttgaa tgaaaatatt ttgaaatgcc ttaactgtct aagattttac 720
tttaaatatt ggaacttta aagaagcatt atagggaaca gccttttttc atgcacttat 780
ggtaaataac tataaaaaca aatgaattac aataaattta taattcatga caactgaatt 840
tgggaaaggt aatagttaag tgtttttcca ctaaattact ttttttctaa tcagtgtgaa 900
gtgacacagg aaagtaaaat tgtcccttat aaataggctt tattttaaat gtcaaagaaa 960
attaaagaat ttcacaataa atgtacatgt tgttgctgtt gacaagtgtt gtatgtgaag 1020
gtgatttcat ctga 1034